

Dangerous nuclear whispers

Voices within the Obama administration threaten to undermine non-proliferation efforts. They should be ignored.

In April, US President Barack Obama told a crowd in Prague of his hope for a world free of nuclear weapons. Achieving such a vision would take “patience and persistence”, he said. But we “must ignore the voices who tell us that the world cannot change”.

Unfortunately, some of those voices belong to powerful figures within Obama's own administration. Last month, the non-proliferation news site Global Security Newswire reported that the defence secretary, Robert Gates, has been quietly lobbying for the reinstatement of a programme to build a new generation of ‘reliable’ nuclear warheads. The same report suggests he has at least tentative support from two key cabinet members: Secretary of State Hillary Clinton and energy secretary Steven Chu. However, pursuing this programme would do nothing to improve the reliability of the US nuclear arsenal. Instead, it would seriously damage Obama's efforts to reduce the threat posed by the world's most deadly weapons.

Since 1992, the United States has held a voluntary moratorium on nuclear testing. During that period, scientists at the nation's nuclear-weapons laboratories have used computer simulations and non-nuclear tests to ensure that the existing weapons are safe and dependable.

During the administration of President George W. Bush, however, some weapons scientists sought to move beyond this post-cold-war caretaker role, and pushed for the development of low-yield and earth-penetrating nuclear weapons that could be used against conventional military targets. Congress halted those projects for fear they would rekindle the arms race. So the scientists came up with a supposedly more benign warhead concept. Dubbed the Reliable Replacement Warhead, it would require less maintenance than existing weapons, and would last longer. One argument was that the plutonium in current-generation weapons was degrading because of its own radioactivity, and that this, over time, would make the devices too unreliable to use. Another, which Gates repeated in a speech last October, was that the new weapons would not require testing. Ulti-

mately, he asserted, such devices would allow the United States to further lower its nuclear stockpile.

Yet such arguments are spurious. Studies by the weapons labs themselves show that the nuclear material within existing devices will last for decades (see *Nature* **444**, 660–661; 2006). Experienced nuclear-weapons scientists believe that it would be irresponsible to deploy a warhead without testing it first. And there is no reason that the stockpile could not be reduced without building a new nuclear warhead.

Gates is now leading a major administration review of the entire nuclear-weapons complex, including the reliable-warhead proposal. That review comes at a crucial time for its non-proliferation agenda. Obama is currently pursuing Senate ratification of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), an international prohibition of nuclear-weapons testing. Next year, his administration will also take part in an international review of the Non-Proliferation Treaty, the main international tool to limit the spread of nuclear weapons. For the United States to be developing a new warhead during this period would look to other nations like rank hypocrisy. Moreover, the replacement programme's very conceit, that existing warheads may not be reliable for much longer, will probably fuel conservative resistance to ratification of the CTBT.

Hopefully, the nuclear review will decide against recommending any sort of replacement programme, ‘reliable’ or otherwise. But if it does not, then Obama should have the courage to reject the plan. The US nuclear stockpile is more than adequate to defend the nation's territory and that of its allies for decades to come. If Obama truly wishes to lead the world in nuclear disarmament, then he should do so with the warheads the nation already has. ■

“There is no reason that the stockpile could not be reduced without building a new nuclear warhead.”

Cash costs

Massive funding for Pakistan's ailing universities holds many lessons for other developing nations.

Eight years ago, a task force advising Pakistan's former military ruler, General Pervez Musharraf, laid out a bold plan to revitalize the country's moribund research system: initiate a fivefold increase in public funding for universities, with a special emphasis on science, technology and engineering. The proposal was a radical departure from conventional wisdom on the economics of developing nations, which favours incremental investments. Sudden surges of cash are held to be dangerous in poorer countries, which often lack the institutions

or the calibre of people required to make the most of such a windfall, and the money can easily be wasted or fall prey to corruption.

Nonetheless, Musharraf agreed to the proposal. The reforms began in 2003. And the results, which have now earned a qualified thumbs-up from a group of experts in science and education policy (see page 38), offer some valuable lessons for other developing nations.

First, conventional wisdom isn't always right. Despite early doubts that Musharraf's autocratic regime could allocate the new funds effectively, the experts cite initiatives such as a free national digital library and high-speed Internet access for universities as examples of success, as well as new scholarships enabling more than 2,000 students to study abroad for PhDs — with incentives to return to Pakistan afterwards. And they acknowledge that the years of reform have coincided with increases in the number of Pakistani authors publishing in research

journals, especially in mathematics and engineering, as well as boosting the impact of their research outside Pakistan.

Second, human capital matters. One concern raised by the report published in this issue is that the 3,500 candidates for Pakistan's new domestic PhD programmes have had lower qualifications than the candidates going abroad. But that is a situation that should correct itself over time as Pakistan's schools improve. For the time being, the more important point is that Pakistan has opened up the chance of a research degree to many more people than in the past — including those who do not have wealthy families, or access to influential people, or good skills in European languages. Harnessing those reserves of talent is an integral part of any nation's development.

Finally, accountability is essential. This was not a priority for the architects of Pakistan's educational reform, partly because they were working for an autocratic regime, and partly because they were in too much of a hurry. The government seemed to be living on borrowed time, Musharraf's science adviser, Atta-ur-Rahman, has recalled. On the one hand, politicians, judges and lawyers were pressing for a return to democracy; on the other, the influence of the Pakistani Taliban was

increasing. Suicide bombers twice tried to assassinate Musharraf — once by blowing up his motorcade as he returned from making a speech to scientists. If the reformers didn't get their programme in place quickly, they feared they might not get it in place at all.

The result, however, is that the body created to implement the reforms, the Higher Education Commission, has operated with minimal oversight by academics, parliamentarians or anyone else. There has been some waste, although no one has yet accused the commission of egregious abuses of power. But it has exhibited blind spots that an outside influence might have corrected — notably a total lack of investment in the social sciences and policy research, disciplines that encourage the asking of questions that autocratic regimes frequently dislike answering.

This must change. Pakistan is no longer a dictatorship. The elected government, under President Asif Ali Zardari, has expressed cautious support for continuing Musharraf's education reforms. It therefore has an opportunity to build on their successes and correct their shortcomings — starting with an independent review of the commission's performance. ■

US visa nightmares

Barriers faced by foreign scientists seeking entry to the United States do more harm than good.

Over the years, the United States has benefited enormously from its ability to attract the most creative scientific minds from around the globe. Increasingly, however, scientists, postdocs and students are turning elsewhere, frustrated by the barriers to gaining entry that sprang up in the wake of the terrorist attacks of 11 September 2001. In its current incarnation, the US visa application process not only presents a bewildering tangle of directives, prerequisites and requirements, but has also forced some applicants to wait up to a year for their visa to be approved — often for no apparent reason.

It is true that waiting times are improving. In late May, in response to mounting protests from the scientific community and elsewhere, the Department of State, which oversees the processing of student and exchange visas, added staff and resources, and implemented new procedures to cut delays. The department established a 10-day deadline for most applications that require 'administrative processing', a particular security check required when the applicant is from certain nations or does certain work. Among those getting fast-tracked is the Visa Mantis security check, the type that most often affects scientists.

'Routine' approvals should take only a few days, and never longer than a month, the state department says (see page 131). In an effort to pinpoint and resolve other visa-related problems, the department meets regularly with an inter-agency group that includes the Department of Homeland Security, which handles employment and other types of immigration visas. A meeting is scheduled for this week.

But reducing waiting times for visa approvals fixes only one part of an inept and dysfunctional system. Consider, for example, the impenetrable snarl of bureaucratic requirements that an individual must meet before even applying. A list on the state department's website

of required documentation for all students — not just nuclear physicists, or scientists from countries that sponsor terrorism — is an eye-glazing jumble of acronyms, abbreviations and conditions. It shouldn't be this complicated.

Then there are the requirements and conditions of various visas. Almost all applicants must prove that they plan to return to their home country when the visa comes to an end. Those who receive government funding, or whose speciality is on a skills list negotiated by their home country, must return home for two years once the visa expires. Meanwhile, student applicants, as well as postdocs or scientists applying for an exchange visa, must prove that they can cover their expenses.

Employment-visa applicants face their own woes. There is an annual cap of 65,000 such visas for individuals being employed by private- or public-sector companies, plus another 20,000 for individuals with at least a master's degree from a US institution. Applications are accepted each year from 1 April. If the cap is reached before an individual's application is processed, his or her only recourse is to reapply the next year — and risk losing the job.

Capitol Hill staffers say that several legislators are keen to simplify or even throw out many of these rules, but add that the visa troubles are part of a much larger immigration-reform problem. Given the many other issues facing Congress at the moment, from health-care reform to financial reform, action on immigration seems unlikely before next year at the earliest. And even then, months of debate will be required before any new legislation passes — legislation that may or may not address the visa challenges.

So in the meantime, US agencies should act quickly to streamline the application process as far as possible without legislation. All caution should not be abandoned, but at the same time it should not be so difficult for a scientist or student to seek entry to the country for scientific purposes that have no link with terrorism. The United States cannot continue to bar the door to some of the very people it needs most. ■

RESEARCH HIGHLIGHTS

EXOPLANETS

Explaining the eccentricities

Asrophys. J. **702**, 716–723 (2009)

The near-circular orbits followed by the planets in the Solar System are probably not the norm. Many planets travelling around other stars appear to follow an oblong or 'eccentric' path.

In some cases, that apparent eccentricity may be caused by the presence of a second, undetected planet. Timothy Rodigas and Philip Hinz at the University of Arizona in Tucson modelled a range of solar systems. They found that about 4% of the time, a hidden planet far from the star could create the illusion that the closer planet was in a highly eccentric orbit. The effect was much more pronounced for planets with moderately oblong orbits.

The team says that their study could help to explain the apparent abundance of eccentric exoplanets. It could also help researchers to find hidden planets.

BIOLOGY

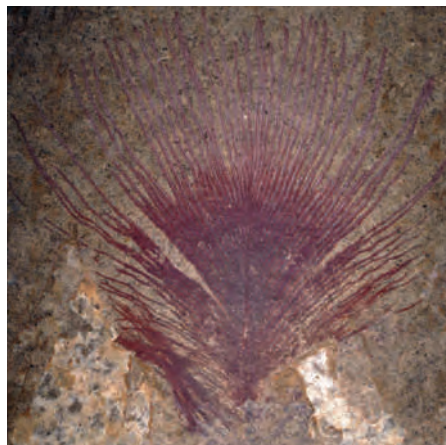
A colourful past

Biol. Lett. doi:10.1098/rsbl.2009.0524 (2009)

Reconstructions of extinct species often attribute garish colours to animals on the basis of nothing more than artistic licence. Now Derek Briggs of Yale University and his colleagues have discovered evidence of preserved colour in fossil feathers.

Using a scanning electron microscope, they found neatly organized pigment structures called melanosomes in fossils such as the one pictured below from the Messel Shale in Germany. Owing to the loss of surrounding keratin, which is involved in optical scattering, they could not determine the feathers' exact hue.

But comparison with the melanosomes of modern birds suggest black feathers with a iridescent blue, green or copper sheen.



J. VINTNER

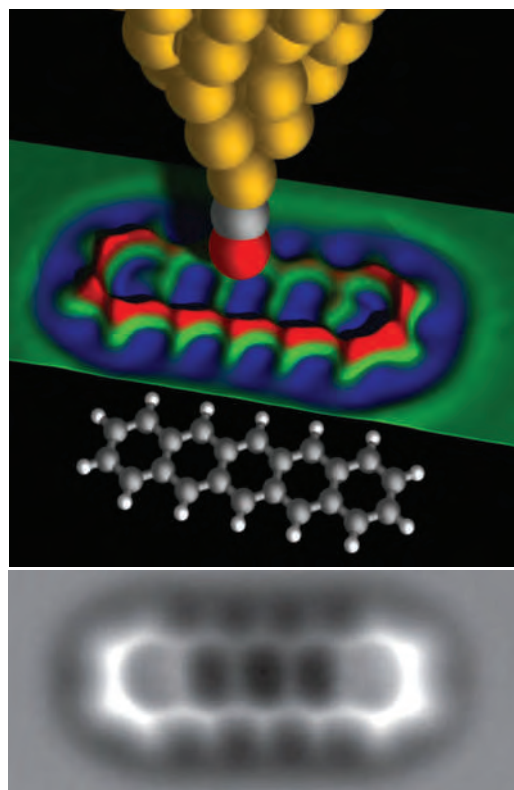
Seeing the honeycomb

Science **325**, 1110–1114 (2009)

Behold pentacene (pictured right), a five-ringed hydrocarbon molecule shown with remarkable clarity thanks to an atomic force microscope with a crucial modification.

Leo Gross and his colleagues at IBM Research in Zurich, Switzerland, placed a carbon monoxide molecule on the tip of the microscope's sensitive detector. The tip scans over the surface and is deflected by its interaction with the atoms in the sample.

Choosing the right molecule for the tip enabled repulsion due to the electronic exclusion principle to dominate over the blurring caused by attractive Van der Waals and electrostatic forces. The improved contrast allowed the researchers to resolve the positions of individual molecules and the bonds between them.



IBM RESEARCH, ZÜRICH

IMMUNOLOGY

Killer fat

Immunity **31**, 232–244 (2009)

Lipid droplets in a type of immune cell may have a key role in triggering attacks on infected and damaged cells.

Dendritic cells engulf pathogens, chop up their components and display the resulting antigens on their surfaces using the major histocompatibility complex class I (MHC I) molecules. These antigens activate cytotoxic T cells, which seek and destroy infected cells.

Pierre Guernonprez at Rockefeller University in New York and his colleagues found that an enzyme residing on lipid droplet membranes is essential to this pathway. When the researchers eliminated the enzyme in mice, they found defects in the cells' lipid droplets and in their ability to display antigens through the MHC I pathway. The manipulation did not affect the MHC II pathway.

PAIN

Deep, deep in your head

Neuron **63**, 533–543 (2009)

The placebo effect is more hard-wired into the brain than previously thought.

For decades, scientists have known that sham treatments can relieve pain, and studies

have identified natural opioid pathways as important players. But because the placebo effect seems to be linked to expectation, some have assumed that it occurs mainly through 'higher' brain structures associated with consciousness.

Falk Eippert at the University Medical Center Hamburg-Eppendorf in Germany and his team imaged the brains of volunteers given a sham ointment to relieve a mild burning pain. Half of them had been treated with naloxone, a chemical that blocks opioid signalling. The researchers found that placebo-related brain activity normally occurs in both the prefrontal cortex and more hard-wired areas, such as the amygdala, hypothalamus and parts of the brainstem.

CHEMISTRY

Bacterial factories

Biotechnol. Bioeng. doi:10.1002/bit.22502 (2009)

Researchers have engineered bacteria to produce useful levels of putrescine, a key ingredient in the production of plastics and industrial chemicals. Putrescine — a natural breakdown product of amino acids that contributes to the characteristic odour of rotting flesh — has historically been manufactured from petrochemicals through environmentally harsh methods.

To develop a more sustainable process, Sang Yup Lee and his colleagues at the Korea Advanced Institute of Science and Technology in Daejeon, Republic of Korea, boosted putrescine production in *Escherichia coli*. Their method involved inactivating the metabolic pathways that use and degrade putrescine, and amplifying the production of an enzyme that makes it from its precursor, ornithine. In high-cell-density cultures, the bacteria can produce up to 24.2 grams of putrescine per litre in about 32 hours.

CANCER BIOLOGY

Cilia's dual role

Nature Med. doi:10.1038/nm.2011; doi:10.1038/nm.2020 (2009)

Cilia, the thin filaments that protrude from many mammalian cells, can both inhibit and exacerbate tumour formation in mice.

Cilia are essential for proper functioning of the hedgehog signalling pathway, which has been found to go haywire in various cancers. Jeremy Reiter of the University of California, San Francisco, and his colleagues deleted genes for cilia formation in two mouse models of skin cancer, each carrying a mutation in the hedgehog pathway. Tumours did not grow in mice with one mutation, but were accelerated in mice carrying the other.

Arturo Alvarez-Buylla, also at the University of California, San Francisco, and his colleagues found similar results modelling brain cancer in mice. In addition, they showed that some types of human brain cancer have cilia, whereas others do not, suggesting that cilia could aid in diagnosing cancer type.

GENETICS

Y-rated

Curr. Biol. doi:10.1016/j.cub.2009.07.032 (2009)

DNA sequencing of the human Y chromosome suggests that humans carry 100–200 genetic mutations not seen in their parents. This direct measurement of the human mutation rate should help researchers to refine evolutionary dating and better understand the source of genetic disease.

Yali Xue and Chris Tyler-Smith of Wellcome Trust Sanger Institute in Hinxton, UK, and their colleagues sequenced the Y chromosome of two men in China separated by 13 generations. Four mutations had cropped up during that time. Extrapolating out to the 6-billion-odd base pairs of the complete human genome, that translates to roughly one mutation per 30 million base pairs per generation on average.

For a longer story on this research, see <http://tinyurl.com/nv9u59>

MICROBIOLOGY

Resistance is futile

Clin. Infect. Dis. **49**, 869–875 (2009)

Antibiotic resistance in pathogens can be reversed by rolling back the total amount of the drug consumed, but it increases as soon as consumption rises again.

During a seven-month period in 2001–02, the use of ciprofloxacin, a popular antibiotic, was restricted in Israel because supplies were being stockpiled in case of a bioterrorist attack. Michal Chowers and her colleagues at Tel Aviv University in Israel measured the ciprofloxacin sensitivity of the bacterium *Escherichia coli* from urine samples before, during and after this period.

A near 50% drop in ciprofloxacin use reduced the percentage of samples containing resistant bacteria from 12% to 9%. But resistance surged as soon as people started popping the pills again.



C. KREBS / CORBIS

FOOD CHEMISTRY

Bee-devilled by corn syrup

J. Agric. Food Chem. **57**, 7369–7376 (2009)

Dangerous levels of toxins that can form in high-fructose corn syrup and endanger domesticated honeybees can now be easily predicted and eliminated.

Blaise LeBlanc, while at the US Department of Agriculture in Tucson, Arizona, and his colleagues measured the rate of formation of hydroxymethylfurfural (HMF), a heat-generated contaminant found in high-fructose corn syrup, honey and other foods. Over a 35-day period, they tracked HMF content in several brands of syrup stored at different temperatures. They also established the concentration at which HMF becomes toxic to caged bees.

Although temperatures of around 32 °C had little effect, those of 40 °C and above caused the HMF content to rise markedly. The rate of the increase can be precisely predicted from pH and other variables, the scientists say. HMF formation was lower in syrups of higher pH, and adding bases to syrup reduced levels of the toxin.

JOURNAL CLUB

Richard Bennett
Brown University, Providence,
Rhode Island

A microbiologist wonders what turns us on.

An Internet search for the words 'pheromone attractant' pulls up products ranging from human aphrodisiacs to control measures for the Colorado potato beetle.

But sexual chemistry is not only important to humans and beetles, it is also relevant to many fungi. Fungal peptide pheromones are often released by one mating type to attract a partner of the opposite sex, thereby initiating the programme of sexual differentiation. This signalling is often highly specific so that pheromones attract only potential partners and not unwanted suitors.

Work by Joseph Heitman and his colleagues at Duke University in Durham, North Carolina, provides a new spin on pheromone signalling in fungi (Y.-P. Hsueh *et al.* *EMBO J.* **28**, 1220–1233; 2009). While studying the fungal pathogen *Cryptococcus neoformans*, the authors became curious about the function of an uncharacterized pheromone-receptor-like gene.

It turns out that this gene, *CPR2*, encodes a constitutively active receptor that stimulates downstream mating events in both the presence and absence of pheromones. During sexual differentiation, expression of *CPR2* is upregulated and supplements the activity of conventional pheromone receptors. A single amino-acid substitution in the Cpr2 protein, in a transmembrane domain that is highly conserved among pheromone receptors, was shown to be responsible for constitutive signalling activity.

This demonstrates that the sexual lifestyles of unicellular organisms can be much more complicated than they first seem. Furthermore, constitutively active receptors have been implicated in many signal-transduction processes in mammalian cells. It remains to be seen whether sexual activity in more complex organisms also involves signalling components that are continuously turned on.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS BRIEFING

● POLICY

Energy strategy: The Indian government has approved a **national trading scheme** for carbon credits and energy-efficiency certificates that it claims could be worth more than 750 billion rupees (US\$15 billion) by 2015. The National Mission on Enhanced Energy Efficiency, one of eight proposals in **India's** climate-change strategy, is projected to reduce the country's energy consumption by 5% a year, and cut about 100 million tonnes of carbon dioxide every year from its current annual emissions of 3 billion tonnes.

Climate resolution: In an unexpected announcement on 27 August, **China's** top legislative body approved a resolution calling for new laws to target energy saving and emissions reductions in the fight against **climate change**. The surprise move came at the close of a four-day session of the Standing Committee of the National People's Congress. It was not originally on the meeting's agenda and, state media said, shows that measures to tackle climate change are moving closer to the heart of government policy.

Flu pandemic: H1N1 flu vaccines will not be ready in the **United States** until mid-October, the US Centers for Disease Control and Prevention reiterated last week — despite a call from the President's Council of Advisors on Science and Technology to accelerate vaccine production to mid-September. But Harold Varmus, co-chair of the 21-strong advisory panel, praised the federal government's response to the virus as “truly impressive”.

● BUSINESS

Gas deal: Australia's environment ministry approved a Aus\$50-billion (US\$42-billion) development of the country's Gorgon **natural gas**

NUCLEAR ROW RESURFACES

Ahead of national elections on 27 September, a dispute flared up last week on how Germany will dispose of its nuclear waste. Documents published by a local anti-nuclear group show that in 1983 Helmut Kohl's government altered a scientific report expressing concern over the Gorleben salt dome (pictured) — chosen in the 1970s to become the national long-term waste-storage facility. A moratorium on exploration of the site is due to expire next year. Chancellor Angela Merkel wants to retain Gorleben; but opposition party — and anti-nuclear — environment minister Sigmar Gabriel said it was “dead” and urged research into alternative sites.

fields off northwest Australia. Energy companies Chevron, ExxonMobil and Shell expect to produce 15 million tonnes of liquefied natural gas a year from the project (about 8% of current global trade), much of it destined for China. Conservation groups worry that the facilities could damage the environment of Barrow Island, a local nature reserve.

Blockbuster drug: AstraZeneca has announced that its blood-thinning drug candidate ticagrelor (Brilinta) outperformed one of the world's best-selling drugs, clopidogrel (Plavix), in a head-to-head trial. In a study involving more than 18,000 patients, ticagrelor was better at reducing heart attack and stroke deaths than clopidogrel (L. Wallentin *et al.* *N. Engl. J. Med.* doi:10.1056/NEJMoa0904327; 2009). Analysts say that the drug would enter a market worth around US\$9 billion.

Acquisition: Irish drug maker **Warner Chilcott**, based in Ardee, has announced plans to buy the prescription-drug business of **Procter & Gamble**,



based in Cincinnati, Ohio, for US\$3.1 billion, all of which will be borrowed. Six banks are fronting \$4 billion to bankroll the acquisition, including \$1 billion to refinance Warner Chilcott's current debt. The transaction, expected to close by the end of the year, would be the largest leveraged loan for an acquisition in 2009, portending the return of a looser lending environment.

Green investment: Government officials in **Norway** announced on 31 August that the country has redirected about US\$4 billion of its \$400-billion sovereign-wealth fund into environmentally friendly companies. The fund invests the country's oil and gas revenues and owns around 1% of global stocks. The change in investment strategy was decided earlier in the year, and includes \$1.2 billion for 232 Indian companies.

Research productivity: European pharmaceutical companies offer a better bang for the buck than their US counterparts, according to a re-analysis of more than 20 years of data on **drug**

SOUND BITES

“Hydrogen fuel cells are losers.”

Burton Richter, Stanford University, California

Speaking on 26 August at IBM's Almaden Institute the Nobel laureate said that today's expensive and short-lived fuel cells must to go back to the R&D lab.

(Greentech Media)

R. JENSEN/DPA

discovery. A 2006 investigation (H. G. Grabowski and Y. R. Wang *Health Aff.* 25, 452–460; 2006) of all novel drugs developed worldwide from 1982 to 2003 concluded that the United States was most productive. But when Donald Light, a health researcher at the University of Medicine and Dentistry of New Jersey in Newark, controlled for the size of companies' investment in R&D, he found that, dollar for dollar, Europe brought more new treatments to market (D. W. Light *Health Aff.* doi:10.1377/hlthaff.28.5.w969; 2009).

RESEARCH

Contract break: The US Department of Veterans Affairs has cancelled a five-year, \$75-million contract with the **University of Texas** to research **Gulf War syndrome**. The department cited "persistent noncompliance and numerous performance deficiencies" as justification for ending the research contract after two years. The university said it "strongly disagreed" with the department's assessment. A July report from the department's inspectorate had accused the university's Southwestern Medical Center at Dallas of breaking the terms of the contract, failing to send bills on time and unilaterally changing an informed consent form after it had been agreed.

Geoengineering: Britain's national academy of science, the Royal Society, has released its first report analysing **schemes to alter Earth's climate** on a global scale. The 1 September

NUMBER CRUNCH

\$67 billion

Annual amount sought by Africa from developed nations to mitigate the effects of climate change. African leaders met in Libya this week to firm up the draft proposal.

report looked favourably on seeding the stratosphere with sulphates to reflect back sunlight, and on filling agricultural land with minerals to absorb carbon dioxide, but cautioned that even preliminary research projects would need international pre-review. For more, see page 19.

Climate change: The global cost of **adapting to global warming** has been grossly underestimated, according to a report released on 27 August by the International Institute for Environment and Development in London. It could be at least 2–3 times more than the 2007 estimate from the United Nations Framework Convention

on Climate Change of between US\$49 billion and \$171 billion per year, the study's authors say. For more, see page 24.

EVENTS

Doctorate scandal: Prosecutors in Germany have made public their ongoing investigation into academics who may have taken **bribes to award doctoral degrees** to students. Professors from a number of universities are suspected of accepting payments of €2,000–5,000 (US\$2,870–7,175) to supervise PhDs — and to ensure that passes were awarded to students that might not have made the grade. The director of a Cologne-based academic consulting firm that allegedly connected would-be PhDs with bribable professors was jailed last year, along with an unnamed law professor from the University of Hanover.



Japan election: The Democratic Party of Japan (DPJ) won a landslide election on 30 August, wresting power from the Liberal Democratic Party (LDP), which had ruled almost continuously since 1955. The DPJ, which is led by **Yukio Hatoyama** (pictured), has promised to increase the science budget and replace the country's highest science-policy body, the Council for Science and Technology Policy, with a 'strategy office' that would

THE WEEK AHEAD

6–10 SEPTEMBER

The 16th Congress of the International Society of Developmental Biologists meets in Edinburgh, UK
www.isdb2009.com

8 SEPTEMBER

The US Congress returns from its summer recess. Later this month, the Senate is expected to introduce for discussion its version of a cap-and-trade bill for carbon emissions.

promote science more actively. It has also called for a cut in greenhouse-gas emissions to more than 25% below 1990 levels by 2020, a much bigger reduction than the LDP's 8% proposal (see *Nature* 460, 938; 2009).

Lunar losses: India formally abandoned its first lunar orbiter on 30 August, after scientists at the Indian Space Research Organisation (ISRO) abruptly lost radio contact with the probe. **Chandrayaan-1**, launched last year to map the Moon, ended its mission 14 months early, but the ISRO said it had met most of its scientific objectives. Meanwhile, **NASA's Lunar Crater Observation and Sensing Satellite (LCROSS)** accidentally burned up most of its spare fuel on 22 August. Mission managers say it remains on track to smash into the Moon on 9 October, in the hope of kicking up evidence of ice.

D. KUROKAWA/EPA/CORBIS

BUSINESS WATCH

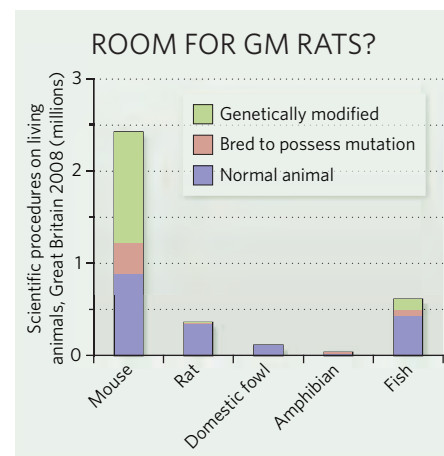
Chemical reagent company Sigma-Aldrich is to spin off a business that will create genetically modified (GM) rats. Sigma Advanced Genetic Engineering Labs, based in St Louis, Missouri, should start taking orders by October, according to Sigma representatives. Early releases will include knockouts for the genes *APOE1*, implicated in atherosclerosis and Alzheimer's, and *DISC1*, associated with schizophrenia.

The market for GM animals is expected to grow by 12% a year over the next 3 years, says Eric Ostertag of Transposagen Biopharmaceuticals, based in Lexington, Kentucky, one of the world's only suppliers of GM rats. But it is dominated by knockout mice; in Britain last year, they

outnumbered GM rats by 1.2 million to 6,000 (see chart; figures exclude Northern Ireland).

Rats are seen as superior to mice as models for several human diseases, but are harder to manipulate genetically using embryonic stem-cell methods. Sigma-Aldrich uses zinc finger nucleases — a technology licensed from California-based Sangamo Biosciences — to delete specific genes directly in the rat embryo (see A. M. Geurts *et al. Science* 325, 433; 2009). The firm will launch its first catalogue of animals in October at the Society for Neuroscience meeting in Chicago, Illinois.

For a longer version of this story, see
<http://tinyurl.com/m78fw7>



SOURCE: UK HOME OFFICE

Climate-control plans scrutinized

The Royal Society reviews options for fighting global warming with geoengineering.

As atmospheric levels of carbon dioxide soar and political efforts to control emissions stagnate, one scientific academy says that it is time to consider radical intervention.

On 1 September, the Royal Society, Britain's premier scientific organization, released its first analysis of a host of controversial methods for intentionally altering Earth's climate. Such approaches, known broadly as geoengineering, could slow or halt climate change by either restricting the amount of sunlight heating Earth's surface or reducing levels of carbon dioxide in the atmosphere.

But they could also have disastrous side effects, such as stifling the Asian monsoons or altering the oceans' oxygenation or pH. For these reasons, the society is calling for international regimes to review even preliminary research projects. Further work, it says, must also go into developing rules for how and when geoengineering might be used.

Because the science and policy surrounding geoengineering is so complex, any potential scheme would take decades to put into practice, says John Shepherd, an Earth scientist at the University of Southampton, UK, who chaired the report. "These things may help us get out of a fix later in the century," he says. "But we have to do research now."

Geoengineering proposals have been around for decades, but most have been dismissed out of hand because they often seem outlandish, costly or downright dangerous. Some researchers see the Royal Society report as the first real indication that geoengineering is being taken seriously. "This is actually quite a big deal," says David Keith, an environmental engineer at the University of Calgary in Alberta, Canada, and one of a dozen authors on the study. The report is "the first major national academy report solely devoted to this topic," he says.

Shepherd stresses that the Royal Society wasn't trying to pick winners and losers among possible geoengineering approaches, but the group did weigh in with which concepts they consider the most promising for future study (see 'Thinking big'). Some popular ideas received low marks. The idea that painting roofs white could cool urban areas — proposed among others in a 27 August geoengineering report by the UK Institution of Mechanical



Geoengineering ideas include using sea salt to brighten clouds.

Engineers — would produce a near-negligible change in Earth's overall temperature, the Royal Society argued. Also low on the list was using iron to increase carbon-absorbing algal blooms in the ocean. Such algal outbreaks would absorb relatively little carbon, the new study warns, and would consume vast amounts of oxygen, potentially leading to oceanic 'dead zones'.

Other ideas were viewed more favourably. Artificial weathering, the acceleration of geological carbon-absorbing processes, was among the most promising ways to capture carbon, according to the panel. Filling agricultural land with carbonate and silicate compounds

could turn fields into carbon sponges, absorbing massive amounts of carbon dioxide. And seeding the stratosphere with sulphate aerosols — a process that would increase the atmosphere's reflectivity — could mimic the immediate cooling effects of major volcanic eruptions. The sulphate strategy has the advantage of being a quick way to cool the planet, Shepherd says. But to keep it cool, "you have to keep it going for many decades or possibly centuries," he says. "There is no silver bullet."

Dangerous distraction

Most of the top-ranked strategies do have potentially catastrophic side effects. Some models suggest

that putting sulphates into the stratosphere could degrade the ozone layer or alter the monsoons — affecting the livelihoods of billions of people. Using silicate or carbonate compounds, meanwhile, could alter the pH of soil or marine ecosystems, depending on how researchers dispose of the material. For these reasons, experiments must undergo stringent ethical review and public discussion, says Michael Oppenheimer, a geoscience and policy researcher at Princeton University in New Jersey who reviewed the study. "Scientists should not take a step in the direction of field experiments until these issues are resolved," he warns.

Panel members acknowledge that the biggest risk from geoengineering may be that it distracts from the main accepted way of stopping climate change: reducing the amount of greenhouse gases being spewed into the atmosphere. "Geoengineering is already being seen as an easy way out," warns Doug Parr, chief scientist of environmental campaigners Greenpeace UK. Shepherd agrees that geoengineering is no substitute for curbing fossil-fuel use and developing alternative energy sources. "In [the panel's] view, the primary focus has to remain on conventional emissions," he says.

For now, Shepherd thinks that a modest annual expenditure of around £10 million (US\$16 million) should be enough to determine which schemes hold the most promise. "In five to ten years' time," Shepherd predicts, "we'll have an idea whether it's worth spending serious research money on this."

Goeff Brumfiel

Thinking big

Five ideas for engineering the planet.

Stratospheric aerosols Spraying the atmosphere with sulphates is risky, but could cool the planet quickly.

Air capture Man-made scrubbers to remove carbon dioxide directly from the air could work, but are likely to be expensive.

Cloud brightening Spraying sea salt into the sky to brighten clouds would only cool the region below.

Enhanced weathering using silicates or carbonates to absorb CO₂ could work, but more study is needed.

Space reflectors to cool Earth would cost trillions of dollars. The panel recommends keeping space-based research desk-based for now.

G.B.

SPECIAL REPORT

Pandemic flu: from the front lines

As the novel H1N1 pandemic flu virus infects people worldwide, researchers in some of the affected countries describe in their own words the scientific and public-health challenges they face.

MEXICO

POPULATION 110 MILLION

Data suggest that Mexico has seen two waves of infection — the first, which peaked in late April, affected the Mexico City area, and the second, broader wave spanned June through August in southern states, including Chiapas, Yucatan and Quintana Roo. To prepare for a potentially larger wave this winter, Mexico is raising public awareness, standardizing timely diagnosis and treatment and reinforcing equipment and management protocols in intensive-care units throughout the country.

To improve surveillance, Mexico has accelerated the upgrading of its public-health laboratory network. The national reference laboratory and 28 states will soon have real-time PCR for running diagnostic tests. This builds on a restructuring of Mexico's national surveillance and reporting systems, which started in 2007.

As Mexico's strategic reserve of antivirals would cover only 1% of the population for community cases and up to 80,000 hospitalized cases, the nation is implementing a central logistics and delivery system to assure their efficient allocation. The country also expects to have 20 million doses of the H1N1 vaccine available by December. As even this would cover only a fraction of the population, the government will prioritize health-care workers, then individuals at risk of severe disease, such as pregnant women and people with chronic underlying illnesses.

Stefano Bertozzi, executive director at the Center for Evaluation Research and Surveys at the National Institute of Public Health in Cuernavaca

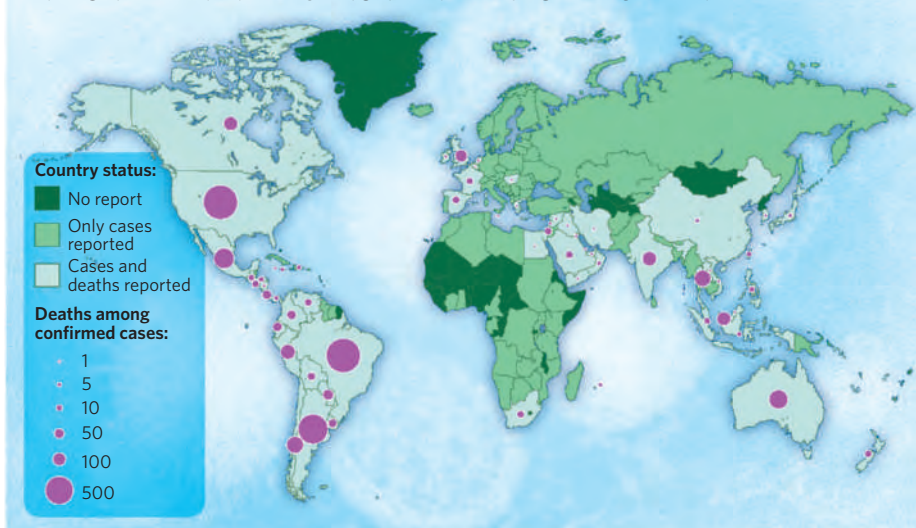
AUSTRALIA

POPULATION 21 MILLION

The timing of the epidemic has differed across the country, which has meant that we needed different public-health measures and messages in individual states. The pandemic virus seems to be outcompeting the seasonal flu viruses. The great majority of flu cases around the country are now pandemic H1N1.

One interesting question is whether this pandemic virus will completely replace any of the seasonal flu strains. If it doesn't, that's going to complicate the production of future seasonal flu vaccines, as we will need a vaccine against four strains instead of the current three. The Australian government has ordered 21 million

REPORTED FATALITIES FROM PANDEMIC INFLUENZA



doses of dedicated pandemic virus vaccine, so if we need two doses per person, that covers half the population. There has been a lot of discussion about who should get it first, and when.

We are seeing similar patterns of disease severity to those reported worldwide, with most cases being mild. But there have been a significant number of cases with severe disease, not just in the at-risk groups, but also in healthy people. Our indigenous population is being hit harder, and we are seeing disproportionate numbers hospitalized with severe disease.

An important message for other countries that have intensive-care facilities is to expect significant pressure on them. There is a need for mechanical ventilators, and we have seen heavy use of scarce extracorporeal membrane oxygenation units.

Anne Kelso, director of the World Health Organization Collaborating Centre for Reference and Research on Influenza in Melbourne

JAPAN

POPULATION 128 MILLION

Japan stopped counting cases on 25 July and launched a new cluster surveillance system that is directly in the hands of the health ministry. Our Infectious Disease Surveillance Center no longer has any disease data feed, making it difficult to analyse epidemiological trends or disease burden. But we have received hundreds of

reports through routine sentinel-based surveillance of clusters of disease from many regions and big cities, so there is extensive spread.

The demand on public-health services to report and investigate all cluster cases is overwhelming public-health staff and leading to a breakdown in the normal public-health diagnostic service in local laboratories.

With the rising numbers of cases we are seeing a corresponding increase in deaths. As elsewhere, it is younger people who are affected with more severe disease requiring hospitalization, but the overall hospitalization rate is no greater than that of human seasonal influenza. Japan has an ageing population with large numbers of people older than 65, many with at-risk underlying health conditions, but so far pandemic H1N1 seems to be largely sparing the elderly.

The country's pandemic plan was based almost entirely on a severe pandemic of H5N1 avian influenza, which limited medical consultations to just a few hospitals.

The government seems to be relaxed with the low level of epidemic by the less virulent virus since May, and seems to have yet to draw any lessons from the pandemic. As a result, local and regional authorities have now independently started to prepare for the coming flu season.

Masato Tashiro, director of the Department of Viral Diseases and Vaccine Control at the National Institute of Infectious Diseases in Tokyo



Q&A: STEPHEN MINGER
Stem-cell scientist to head
GE-Geron drug screening
collaboration.
www.nature.com/news

KING'S COLLEGE, LONDON

ARGENTINA

POPULATION 40 MILLION

The current epidemiological situation is a generalized spread of the virus throughout the country, although with a marked downward trend in the number of reports of the levels of influenza-like illness. The epidemic started in mid-May in Buenos Aires, and three weeks later spread to the city's larger metropolitan area. Activity peaked on 25 July, with influenza A representing 80% of the circulating respiratory viruses; 65% were H1N1-pandemic confirmed. Very few isolates were H3 and H1 seasonal.

Health systems in Argentina were overloaded because of government advice to people to consult a physician on first signs of flu symptoms such as fever or cough.

The major challenge at the lab level was in diagnosing the first cases produced by a new, unknown virus. Later, the challenge was for lab capacity to meet demand. Information transmitted to the public was not always clear enough, and the mass media had a negative role, including providing contradictory information and producing fear.

Wilma Savy, head of the respiratory virus service at the National Institute for Infectious Diseases in Buenos Aires

VIETNAM

POPULATION 85 MILLION

The first cases in Vietnam were at the end of May, a bit later than in many other parts of Asia, probably because Vietnam does not have a major international airport hub. We are now seeing an increase in disease and a small number of severe cases. Vietnam was a hotspot for H5N1 avian influenza in 2003 and 2004, and the pandemic preparation that resulted from both this and SARS has made a massive difference to the current situation.

Prior to avian flu, few hospital staff had community-acquired pneumonia on their radar; attention was concentrated on malaria, dengue fever and tuberculosis. Now clinicians have a much greater awareness of the need to look out for clusters of respiratory illnesses. There has also been greater interaction and collaboration between clinical and other researchers, and between centres across the country.

Access to vaccines and drugs remains an important issue. There is a global shortage of vaccines, and the rich countries have bought up all the first stocks. This is a really urgent issue; if we can get this right now, then many of the past issues around sharing of samples, data and general openness on emerging infectious diseases will be helped, maybe resolved. If we get it wrong, we will be back to square one. If ever there was a time for the rich world to reach

out and ensure equity of access to drugs and vaccines, it is now.

Jeremy Farrar, Vietnam director for the Wellcome Trust Major Overseas Programme, director of Oxford University's Clinical Research Unit in Ho Chi Minh City, and coordinator of the South East Asia Infectious Disease Clinical Research Network

UNITED STATES

POPULATION 301 MILLION

US health-care systems have been stretched and have no surge capacity. The system cannot handle this pandemic, even if it remains moderate in severity. The same applies to many of the supplies we get. Ask anybody who has tried to order an N95 respirator recently; there aren't any. We recently surveyed a group of world-class pharmacists to identify the essential drugs needed daily to keep patients from dying. They came up with a list of more than 30 — all generics, and most made offshore, mainly in Asia, and China and India in particular. Nobody is thinking what might happen to US or global supply chains when pandemic flu hits these countries, where the primary workforce are the young, who are most affected by the virus.

The United States has a federal programme for vaccine procurement but it is administered at the state level, and the two do not always mesh up. It is still not clear how this vaccine is going to be rolled out, or whether it will be here in time.

I worry most that, given current existing public concerns about vaccines, in the autumn we might see mounting public responses and concerns about pandemic-vaccine safety, and people refusing to be vaccinated. Expect the unexpected over the next six months.

Michael Osterholm, director of the University of Minnesota's Center for Infectious Disease Research and Policy in Minneapolis

INDIA

POPULATION 1.1 BILLION

The virus is now transmitting in city clusters. Large numbers of people are turning up at designated testing facilities, swamping an already stretched surveillance system, so there is little room for monitoring mutations and reassortment. This should be done. One way would be to bring in academic labs outside of the government testing system, but sharing of clinical materials and trust is low.

Deaths have sparked a fair amount of concern and panic. Poor communication of risks by the government and the public-health system is largely to blame.

Even if this pandemic remains moderate, the impact in India is likely to be severe, owing to its high population density, low awareness of

the pandemic and the propensity of the virus to infect the young (50% of Indians are under 25 years of age). Moreover, there is a high load of other infectious diseases as well as chronic conditions, groups that are at higher risks of severe forms of pandemic H1N1 disease. The health-care infrastructure is poor.

Despite this bleak outlook, India has strengths for tackling the virus, including that the government has pandemic plans in hand, and that we have a vibrant generic-pharmaceutical industry as well as a decent capacity for manufacturing vaccines. There is little clarity, however, as to India's vaccine plans, and the regulatory process is archaic, so it is not even clear whether pandemic vaccine could be rapidly approved for use in the country. The government says it has enough Tamiflu for 3 million people.

Shahid Jameel, head of the virology group at the International Centre for Genetic Engineering and Biotechnology in New Delhi

SUB-SAHARAN AFRICA

POPULATION 800 MILLION

H1N1 has not yet been reported in Nigeria, or any of the other sub-Saharan African countries with which we collaborate — Niger, Burkina Faso or the Central African Republic, although the Democratic Republic of Congo has one confirmed case. But surveillance is still very poor, and the virus may well often escape detection. International media attention to the pandemic is probably more than it deserves from an African public health point of view. Any diversion of resources from other important programmes needs to be carefully evaluated for long-term cost-benefit and sustainability.

Systems for lab surveillance and reporting of respiratory illnesses have improved since H5N1, which has hit nine sub-Saharan African countries since it first spread to the continent in 2006. With international support Nigeria, for example, has set up a central national laboratory for human influenza surveillance in Abuja, as well as several decentralized satellite labs.

There is no culture of testing for respiratory viruses, however, and the effort that went into H5N1 control is losing steam. The H5N1 virus was perceived as a major threat to the poultry industry, whereas the disease burden of pandemic flu seems low. Don't expect much mobilization for a virus where most cases are mild. ■

Claude P. Muller, head of the Institute of Immunology at the WHO Collaborative Center for Reference and Research on Measles Infections in Luxembourg

Interviews by Declan Butler

See www.nature.com/swineflu for more on pandemic flu.



W. PHILPOTT/REUTERS

Keeping genes out of terrorists' hands

Gene-synthesis industry at odds over how to screen DNA orders.

A standards war is brewing in the gene-synthesis industry. At stake is the way that the industry screens orders for hazardous toxins and genes, such as pieces of deadly viruses and bacteria. Two competing groups of companies are now proposing different sets of screening standards, and the results could be crucial for global biosecurity.

"If you have a company that persists with a lower standard, you can drag the industry down to a lower level," says lawyer Stephen Maurer of the University of California, Berkeley, who is studying how the industry is developing responsible practices. "Now we have a standards war that is a race to the bottom."

For more than a year a European consortium of companies called the International Association of Synthetic Biology (IASB) based in Heidelberg, Germany, has been drawing up a code of conduct that includes gene-screening standards. Then, at a meeting in San Francisco last month, two of the leading companies — DNA2.0 of Menlo Park, California, and Geneart of Regensburg, Germany — announced that they had formulated a code of conduct that differs in one key respect from the IASB recommendations.

Both codes involve an automated step, in which the genes in a customer's order are compared against those from organisms on lists such as the US Centers for Disease Control and Prevention's 'select agents' list. This step uses computer programs such as the US National Center for Biotechnology Information's Basic Local Alignment Search Tool (BLAST), which searches for similarities between gene sequences.

But although the IASB standard specifies that a human expert will follow up on possible 'hits'

identified in the automated screening step, the DNA2.0/Geneart code ends with the automated screening step. The two firms are now merging their databases of genes of concern. This worries some observers, because it is difficult to translate the list of select-agent organisms into lists of dangerous genes (although a panel convened by the National Academy of Sciences is meeting in Washington DC on 3 September to grapple with this problem). And no one believes that such lists will catch every dangerous gene. For instance, they might not identify genes from harmless organisms that had been modified in some new and deadly way, so many companies use human expertise to review the results.

Lowest common denominator

But human expertise costs money, and competition is fierce in the gene-synthesis field, with roughly 50 dedicated companies fulfilling some 50,000 gene orders per year. And some observers think that the industry as a whole might now adopt the cheaper DNA 2.0/Geneart standard. That concerns Markus Fischer, a member of the IASB board and a managing director of Entelechon of Regensburg.

"The proposal from DNA2.0 and Geneart is a kind of lowest-common-denominator idea," Fischer says. "Simply taking a list of genes, performing a BLAST against them and taking a sort of threshold cut-off and saying everything below that cut-off is not of interest to us is frankly a little bit naive and dangerous."

Claes Gustafsson, vice-president of sales and marketing for DNA2.0, counters that human screening is also not perfect. "There's no way to standardize it," he explains. And as for the incomplete nature of databases of select-agent

genes, "we're just going to deal with the stuff that we know something about," he says. "How do you deal with the unknown? It's outside the scope of science."

He also says that discussions on governance of gene synthesis have been going on among policy experts and governments for many years, with no definitive conclusion. "We decided to standardize everything, make it consistent, and move on," he says.

Other companies have not yet decided where they stand on the issue. In November, the IASB will convene a meeting in Cambridge, Massachusetts, to discuss adopting its code of conduct. But already, some companies are intrigued by the DNA2.0/Geneart proposal.

"Our intention is to review the [DNA2.0/Geneart] process and seriously consider switching to that," says John Mulligan, founder and chief scientific officer of Blue Heron Biotechnology, a gene-synthesis company in Bothell, Washington. His company might, however, still keep some of its manual scrutiny. "What I see may come out of this is a standardized minimum across the industry with some companies choosing to augment that," he says.

Maurer says he hopes that government officials in the United States, the country most concerned about biosecurity, will step in and communicate with industry about its preferred standard. So far, many branches of the government have been involved in working on potential regulations, but none has offered opinions on concrete issues such as screening standards.

"I think if the government expressed an opinion, DNA2.0 would blink," Maurer says. "A little bit of effort now would steer them towards the top of existing practice rather than near the bottom."

Erika Check Hayden

"Now we have a standards war that is a race to the bottom."

Stem-cell projects falter

California's troubled economy has hit the state's ambitious stem-cell research programme, delaying the construction of facilities and disrupting recruitment.

At least three of the dozen groups that received a share of US\$271 million in building grants from the California Institute for Regenerative Medicine (CIRM) won't make the construction deadline of the end of 2010. And efforts to recruit stem-cell researchers have stalled at some institutions, because of the delays in setting up lab space and because of university hiring freezes.

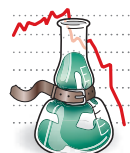
At the University of California, Berkeley, administrators had hoped to recruit 15 principal investigators for a new \$160-million building, including two floors paid for with \$20 million from the CIRM. The building is on track to open in January 2011, "but we have no recruitments under way", says Mark Schlissel, the university's dean of biological sciences. "My real fear is that the University of California system and California will recover slower than the United States."

The CIRM was created by a statewide vote in 2004 to pump \$3 billion into stem-cell research when President George W. Bush had restricted federal funding for work in the field. The initiative called for spending the money over a decade, so to help keep things on track, when the CIRM handed out building grants in May 2008, it required the projects to be completed within two years. That schedule is now in doubt for some.

In San Diego, the delay may stem from a decision by institutions to join forces. The CIRM gave \$43 million to the Sanford Consortium for Regenerative Medicine, which is made up of the historically competing Salk Institute for Biological Studies, the Scripps Research Institute, the Burnham Institute and the University of California, San Diego.

But because the consortium doesn't have a financial track record, officials say that it couldn't get a cost-effective loan to help finance its \$110-million building, and construction stalled. The group is now seeking a University of California guarantee for a \$60 million tax-exempt bond to underwrite construction so the facility can open in 2011. However, the university system is also experiencing its worst-ever financial stress (see *Nature* 460, 441; 2009).

Sanford consortium officials also negotiated with their major donor, South Dakota philanthropist Denny Sanford, to provide more money upfront. The group arranged to receive a promised gift from Sanford in a lump sum of \$7.25 million now instead of \$10 million later.



RECESSION
WATCH

In Novato, the CIRM facility at the Buck Institute for Age Research has also run into delays. The institute was supposed to match \$20.5 million provided by the CIRM. But it couldn't raise the money and is now hoping to get the money from economic stimulus funds from the US National

Institutes of Health. If that fails, the institute plans to seek a tax-exempt bond. The CIRM has given the institute a deadline of March 2010 to break ground or risk losing its grant.

As plans crawl along, university partners are gauging how this might affect recruitment. Lawrence Goldstein, stem-cell research programme director at the University of California, San Diego, says that the roughly 15 stem-cell researchers recruited in recent years are becoming "landlocked". "Our young people need to expand their labs," he says.

In Santa Cruz, the University of California regents last month bailed the local campus out with \$64 million. That will help to pay for a bioscience building to include a stem-cell research floor partly funded by the CIRM. It, too, is behind schedule.

At the University of California, Irvine, administrators called for bids for a new \$61-million facility, which involved reusing a plan from an already constructed building. The competition for work was intense and, because it didn't involve new designs, the university ended up getting an extra, incomplete floor added to its building for no extra cost. Peter Donovan, co-director of the campus stem-cell programme, says he'd like to finish the floor and recruit for it too, but can't. "We committed to five stem-cell faculty hires," he says, "but recruitments now are frozen."

Nevertheless, the CIRM is pushing ahead to try to lure recruits in the tight fiscal environment. On 20 August its governing board approved \$44 million to bring new scientists to the state. ■

Rex Dalton

For more recession stories, see www.nature.com/recessionwatch.

Cost of climate change underestimated

The global cost of adapting to climate change could be 2–3 times higher than previously thought, says a study published on 27 August.

In 2007, the United Nations Framework Convention on Climate Change (UNFCCC), suggested that the annual cost of climate adaptation, starting in 2030, would be between US\$49 billion and \$171 billion. The new report says this estimate does not account for climate change's effects on key sectors such as energy, manufacturing, tourism and ecosystems.

The question of how to finance adaptation will be at the heart of December's negotiations in Copenhagen to forge an international climate deal to replace the Kyoto Protocol when it expires in 2012.

"There are dangers in having apparently low estimates for the cost of adaptation, which would make adaptation seem like a cheap alternative to mitigation," says the study's lead author Martin Parry, a visiting research fellow at the Grantham Institute for Climate

Change at Imperial College London.

The UNFCCC numbers were not intended to be the final word but rather a ball-park figure to get the negotiations rolling, says Sudhir Sharma, manager of financial cooperation and capacity building at the UNFCCC secretariat in Bonn, Germany. "We clearly indicated that this was not an exhaustive study," he says. "Our objective was to kick-start the process of putting numbers on the cost of adaptation so that other groups could pick up the baton and refine them."

The latest work suggests that in the area of water management, the UNFCCC estimate of \$11 billion per year overlooks the expenses of floods and of transporting water from areas of plenty to areas in need. In health, the UNFCCC figure of \$5 billion per year considered changes only in malaria, diarrhoea and malnutrition in developing nations and not the health burden of

climate change in developed nations. The new report also notes that the UNFCCC estimates excluded the costs of protecting ecosystems and the services they provide. On its own this sector could cost well over \$350 billion per year, says Parry.

The report, published by the International Institute for Environment and Development in London and the Grantham Institute, does not provide specific cost estimates, however.

Those may come shortly from other groups. In mid-September, the Economics of Climate Adaptation Working Group, a consortium of researchers, consultants and economists, will release a report on adaptation costs. And on 29 September the World Bank is expected to launch its own study in the area. ■

Anjali Nayar

For more climate coverage, see nature.com/roadtocopenhagen

Budget instructions

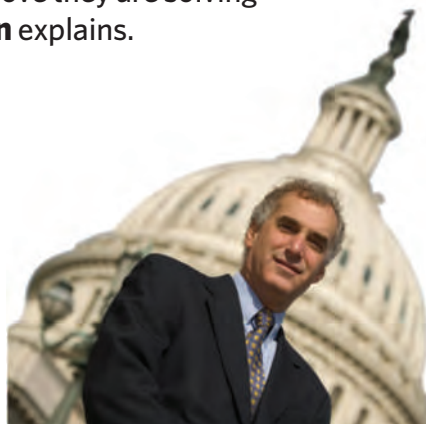
US science agencies may need to prove they are solving national problems. **David Goldston** explains.

Last month, the administration of US President Barack Obama released the first document that hints at how it intends to shape the science and technology budget for fiscal year 2011. If the administration rigorously applies the principles in the document, federal science agencies may need to alter their focus and procedures. For better or for worse, that is making universities' lobbyists jittery.

Although fiscal year 2011 does not begin until 1 October 2010, and the president's proposed budget will not be released until February, drafting has already begun, and this will be the first spending plan that bears the full imprint of the Obama administration. (Fiscal 2009 began before Obama took office, and the 2010 budget was pulled together on a highly abbreviated schedule before all the key administration officials were in place.) So this is the first time that the president's science adviser, John Holdren, and budget director, Peter Orszag, have gone through the annual exercise of jointly issuing a memorandum to guide the agencies as they prepare their budgets for White House review.

What is striking about the 4 August memo is its emphasis on tightly linking research programmes to solving what are pointedly labelled as "practical challenges", its focus on evaluating programme outcomes with data that would be made publicly available, and its call for shifting money away from "lower priority" areas. The document is also notable for what it omits: the term 'basic research' is never used, no mention is made of the president's commitment to double the budgets of several science agencies and no new cross-agency programmes are proposed.

All this, together with the memo's vague directive that agencies should work to improve "the productivity of our research institutions, including our research universities", has been greeted with consternation in universities' Washington offices, and university groups are discussing how to respond. But the memo may win favour with advocates of technology policy, who have complained that the administration is too focused on science and not enough on how to promote innovation in the economy. The document should also bring smiles to academic science-policy experts, who for years have argued that the science budget should be designed more to address specific



PARTY OF ONE

national needs and less to satisfy the interests of researchers and agencies.

The key question, of course, is how the memo will be implemented. Take, for example, biomedical research. Holdren and Orszag define the "practical challenge" for that field as "applying biomedical science and information technology to help Americans live longer, healthier lives while reducing health care costs". Does that mean that the National Institutes of Health will have to make a case that its proposed research will lead in some definable and measurable way to lower health-care costs? Or will it be able to make a generic claim, as it has in the past, that research can reduce health expenses? The memo seems to say that mere assertions will no longer be sufficient, arguing that "agencies have a responsibility to explain how Federal science and technology investments contribute to ... improved health outcomes" and to prove it with data.

Pushing agencies to build their budgets around specific and measurable outcomes that would benefit the public could have a salutary effect — if it doesn't put programmes in a straitjacket. The memo does instruct agencies to "support long-term, visionary thinkers proposing high-risk, high-payoff research" and requires agencies to come up with metrics to determine whether such research is being supported. But an emphasis on specific outcomes, which include "job creation", can push agencies away from riskier projects.

Moreover, no previous administration has been particularly good at figuring out how to evaluate science programmes, especially those that involve fundamental research. The administration of George W. Bush created

something called the Program Assessment Rating Tool to review all programmes, but even Bush officials were cautious in applying it to basic science. In any event, the tool never gained widespread acceptance within the government, and agencies complained that budget decisions often did not take account of the ratings.

A lot of intellectual work will be needed to develop appropriate metrics — some of that is ongoing under the National Science Foundation's Science of Science and Innovation Policy effort — and to limit the burden of gathering data. And a crucial task will always be to decide how quickly an outcome should become evident — inherently an art as much as a science, and a political art at that.

The Obama administration is trying to come up with measures to determine the impact of the American Recovery and Reinvestment Act, which promised short-term job growth, including through science spending. Some university officials were rightly uneasy about the precedent of evaluating science programmes on that basis, but the memo explicitly cites the Recovery Act as a positive precedent for "transparency" and lists job creation as a goal for science and technology programmes, although it does not specify a time period (see *Nature* 458, 563; 2009).

The memo also stresses the link between research and practical results by directing agencies to "empower their scientists to have ongoing contact with people who know what's involved in making and using things", including "cost and competitive factors". The hint of contempt for the ivory tower in the phrasing is surprising, but the idea makes sense for some programmes, especially those that are supposed to yield new energy technologies — another of the memo's "practical challenges". But here, too, the trick will be to get the balance right: an over-emphasis on working with industry or entrepreneurs may limit the scope of research, play into the hands of incumbent market players or create problematic conflicts of interest, especially in the biomedical arena.

The need for balance, however, is not a reason to stick to the status quo. And given the inertia in the funding system, the research enterprise is unlikely to lurch into perilous territory overnight. The administration seems to be sidling into challenging science agencies to think more rigorously about contributing to explicit national goals. But figuring out how to do that, especially if funding is tight, will be a years-long experiment.

David Goldston (partyofonecolumn@gmail.com) is the director of government affairs at the Natural Resources Defense Council in Washington DC. Views expressed are his own.



BATTLEFIELD

Papers suggesting that biotech crops might harm the environment attract a hail of abuse from other scientists. **Emily Waltz** asks if the critics fight fair.

Emma Rosi-Marshall's trouble started on 9 October 2007, the day her paper was published in *Proceedings of the National Academy of Sciences* (PNAS). Rosi-Marshall, a stream ecologist at Loyola University Chicago in Illinois, had spent much of the previous two years studying 12 streams in northern Indiana, where rows of maize (corn), most of it genetically engineered to express insecticidal toxins from the bacterium *Bacillus thuringiensis* (*Bt*), stretch to the horizon in every direction.

Working with colleagues including her former adviser Jennifer Tank at the University of Notre Dame, Indiana, Rosi-Marshall had found that the streams also contain *Bt* maize, in the form of leaves, stalks, cobs and pollen. In laboratory studies, the researchers saw that caddis-fly larvae — herbivorous stream insects in the order trichoptera — fed only on *Bt* maize debris grew half as fast as those that ate debris from conventional maize. And caddis flies fed high concentrations of *Bt* maize pollen died at more than twice the rate of caddis flies fed non-*Bt* pollen. The transgenic maize “may have negative effects on the biota of streams in agricultural areas” the group wrote in its paper, stating in the abstract that “widespread planting of *Bt* crops has unexpected ecosystem-scale consequences”¹.

The backlash started almost immediately. Within two weeks, researchers with vehement objections to the experimental design and conclusions had written to the authors, PNAS and the US National Science Foundation (NSF), Rosi-Marshall's funder. By the end of the month, complaints about the paper had rippled through the research community. By the time Rosi-Marshall attended a National Academy of Sciences (NAS) meeting on genetically modified organisms (GMOs) and wildlife on 5 November 2007,

“She looked hammered”, says Brian Federici, an insect pathologist at the University of California, Riverside, one of those who commented on her work. “I felt really sorry for her. I don't think she realized what she was getting into.”

No one gets into research on genetically modified (GM) crops looking for a quiet life. Those who develop such crops face the wrath of anti-biotech activists who vandalize field trials and send hate mail. But those who, like Rosi-Marshall and her colleagues, suggest that biotech crops might have harmful environmental effects are learning to expect attacks of a different kind. These strikes are launched from within the scientific community and can sometimes be emotional and personal; heated rhetoric that dismisses papers and can even, as in Rosi-Marshall's case, accuse scientists of misconduct. “The response we got — it went through your jugular,” says Rosi-Marshall.

Problem papers

Behind the attacks are scientists who are determined to prevent papers they deem to have scientific flaws from influencing policy-makers. When a paper comes out in which they see problems, they react quickly, criticize the work in public forums, write rebuttal letters, and send them to policy-makers, funding agencies and journal editors. When it comes to topical science that can have an impact on public opinion, “bad science deserves more criticism than your typical peer-reviewed paper”, Federici says.

But some scientists say that this activity may be going beyond what is acceptable in scientific discussions, trampling important research questions and stifling debate. “It makes public discussion very difficult,” says David Schubert, a cell biologist at the Salk Institute in La Jolla, California,

“The response we got — it went through the jugular.”

— Emma Rosi-Marshall

A. SOTIRIOU/GETTY

On a Friday after the paper was published, Federici and plant biotechnologist Alan McHughen, also at the University of California, Riverside, met at a campus bar for a beer after work. “[McHughen] was really annoyed,” says Federici. “I don’t think there’s been another case where I’ve seen him so really ticked off.” Federici says he too was annoyed — Rosi-Marshall’s study was “bad science,” he says, and they feared that activists

Four or other signatories of the McHughen letter went on to publish scathing opinion articles over the next few months. In a March 2008 article⁹ criticizing four papers on biotech crops, Ammann joined forces with Henry Miller, a research fellow at the Hoover Institution in Stanford, California, to ask "Is biotechnology a victim of antiscientific bias in scientific journals?". They called Rosi-Marshall's conclusions "dubious", and said their use of

00004, 00005 = 29471 (October 5, 1991)



Jennifer Tank (left) and Emma Rosi-Marshall study stream ecology.

G. LAMBERTI

evidence “arguably amounts to investigator misconduct”. And in a July 2008 commentary in *Current Science*¹⁰, Shanthu Shantharam, a visiting research scholar at Princeton University in New Jersey said Rosi-Marshall’s “offending” paper “carried a wrong message to farmers and environmentalists”, and that anti-biotech crop activists would use the paper to “hamper the progress of science”.

Rosi-Marshall took the hits hard. “I experienced it in person and in writing,” she says. “These are not the kind of tactics we’re used to in science.” She was a few years out from her PhD, she did not have tenure at Loyola and her first paper in a prominent journal was getting trashed, along with her reputation. “She’s young and was getting picked on,” says Michelle Marvier, a biologist at Santa Clara University in California who attended the NAS November 2007 meeting.

It was at least some comfort to Rosi-Marshall and Tank that e-mails and phone calls of encouragement came pouring in from other scientists. Some of their supporters had observed similar attacks on other biotech crop papers. “The most reassuring thing we learned was that it had happened before and by the exact same people,” says Tank.

What was it about Rosi-Marshall’s paper that prompted such a strong reaction? The wording of the abstract — “widespread planting of *Bt* crops has unexpected ecosystem-scale consequences” — was a particular point of contention. Her critics say that the data do not support such a definitive conclusion. “They absolutely went too far,” says Randy Schekman, editor-in-chief of *PNAS*. Of the half-a-dozen letters received by the journal, most of them protested at this wording, he says. “Why this would have escaped the attention of the referees beats me.”

The authors agree that the wording was unfortunate and in retrospect say that the sentence should have articulated the potential for ecosystem-scale consequences within streams, rather than suggesting that such consequences were observed. “This was an oversight,” says Rosi-Marshall. “But we did not expect that this sentence would, in light of all of the other statements in our paper, elicit the response it

did. We thought the paper would be taken as a whole.”

The study’s methods also came under fire. It is unclear, for example, whether it was the *Bt* toxin itself affecting the caddis flies, or some other difference between *Bt* and non-*Bt* plants. To test this possibility, critics say the caddis flies should have been fed isogenic lines: strains of maize that are genetically identical except for *Bt* genes. The authors say they chose not to use such lines because their nutritional quality would have differed — *Bt* maize has higher concentrations of lignin than non-*Bt* maize, and so is less nutritious. So the authors matched the *Bt* samples with non-*Bt* samples that had similar levels of lignin and other nutrients. “To do otherwise would have resulted in a confounded experiment. Pairing the treatment on the basis of isolines might be standard for agronomic studies, but was inappropriate for an ecological feeding study,” the authors told *Nature* in an e-mail. Rosi-Marshall and her colleagues made this point and other responses to their critics in a correspondence¹¹ published online in *PNAS* the week after McHughen’s and Parrott’s critiques.

It is also unclear how much *Bt* toxin the caddis flies ate. The authors let the insects eat as much as they wanted, as they would in the wild. Critics argue that the authors should have fed the insects known amounts of the toxin in a method called a dose-response study that is routine in toxicity assessments. “The Rosi-Marshall *et al.* paper would have benefited from additional toxicological data,” says Doug Gurian-Sherman, a senior scientist at the Union of Concerned Scientists in Cambridge, Massachusetts, and a former reviewer for the EPA. But the method the authors used “is a widely accepted method, and is generally adequate for a preliminary study of possible toxicity”, he says.

Omitted study

The paper was also accused of omitting contrary findings. In June 2007, four months before Rosi-Marshall’s *PNAS* paper was published, Jillian Pokelsek, a master’s student at Loyola University Chicago working with Rosi-Marshall, presented results from a preliminary field experiment at the annual meeting of the North American Benthological Society in Columbia, South Carolina. The work showed that *Bt* maize pollen did not influence the growth or mortality of filter-feeding caddis flies. The society posted an abstract¹² of the presentation on its website attributing the work to Pokelsek, Rosi-Marshall, Tank, Royer and four other scientists who also authored the *PNAS* paper. It was not mentioning this study that prompted Miller and Ammann’s accusation of misconduct⁹.

The authors defend the omission on the grounds that the data in the meeting presentation were not published or peer-reviewed, and were less reliable than those in the *PNAS* paper. “Field experiments are inherently difficult to control and have lower statistical power to detect significant differences compared with controlled laboratory experiments, thus we included the more controlled and statistically rigorous lab experiments in our paper,” Tank and Rosi-Marshall told *Nature*. Also, the caddis flies in the student presentation belonged to a different family, with different feeding mechanisms to those in the *PNAS* study. Miller’s response: “I don’t want to split hairs,” he says. “If you don’t do appropriate controls or if you draw conclusions that are erroneous, I

“It is critical to assert the right of scientists to question each other’s work.”

— Wayne Parrott

Seeds of discontent

Several scientists say they have been sharply attacked by others in the research community when they have published papers that reflect negatively on biotech crops. In October 2002, for example, David Schubert at the Salk Institute in La Jolla, California, suggested in a Commentary in *Nature Biotechnology* that not enough attention was being paid to the potential unintended molecular effects of inserting genes into plant cells². Almost immediately, he received a barrage of mail from around the world, he says: "I've never received such an obscene response for offering an opinion." Schubert says people complained directly to the Salk Institute, and an administrator called him into his office to say he was jeopardizing funding for his institution. "I've written hundreds of articles — some of them controversial — and never had this kind of response," he says, adding that he has given up trying to have a public discussion about the technology.

One letter¹⁵ critical of Schubert

published in *Nature Biotechnology* and signed by 18 people, admonished: "Good scientists go astray when they leave their area of expertise to offer an opinion when they have not studied the literature." Henry Miller at the Hoover Institution in Stanford, California, also a critic of Rosi-Marshall's paper¹, told *Nature* that "[Schubert] is an accomplished immunologist who has no grasp on agricultural biotech whatsoever." Not true, says Schubert. "The basic technology used to make transgenic plants was invented using bacterial and animal cells, and my lab uses this technology on a daily basis," he says.

In some cases the attacks start before a paper is out. In September 2007, Bruce Tabashnik, an entomologist at the University of Arizona in Tucson, was preparing a paper showing evidence of insect resistance to Bt cotton. He got an e-mail from William Moar, an entomologist at Auburn University, Alabama, warning him that the paper's consequences

would be "devastating". "Your statement ... would be all of the ammunition many special interest groups would need ... Just for a moment think 'monarch butterfly and Bt corn' and the repercussions that surrounded that fiasco," he wrote. Tabashnik's paper was published in *Nature Biotechnology*¹⁶. Moar, who now works for Monsanto, a maker of Bt maize (corn), based in St Louis, Missouri, criticized the paper at conferences and challenged it in Correspondence¹⁷ to the journal saying that the comparisons and conclusions that Tabashnik made were scientifically unsound and based on lab measures, whereas proof of insect resistance must ultimately come from field studies. Tabashnik says: "The rigorous analysis in our paper was based on systematic, objective analysis of all of the relevant data."

One author on Moar's letter was Anthony Shelton, an entomologist at Cornell University in Geneva, New York. He was in action again this year, challenging a review

article by Gabor Lövei, an ecologist at Aarhus University in Denmark, and two co-authors. Lövei's article reviewed laboratory experiments that examined whether crops engineered to kill pests affected the predators and parasites that normally feed on those pests. They found more effects, some negative, some positive, than other reviews had reported. Lövei and his colleagues argued that their method provided a more accurate summary of the literature because it directly examined the data within published papers, rather than relying on authors' conclusions.

Environmental Entomology accepted Lövei's paper¹⁸ but, in January, three months before it was published, Shelton and three colleagues were given a proof by a colleague of one of the authors. Shelton prepared a rebuttal¹⁹ that was published days after Lövei's paper. The six-page critique called the study "negatively biased", "erroneous" and

think that's misconduct." But Ammann says he has a "bad feeling" about the accusation. "Maybe we should have been more careful with the wording."

Scientists who were not involved in the debate over Rosi-Marshall's paper say the results were preliminary and left some questions unanswered, but that overall the data are valuable. "The science is fine as far as I'm concerned," says Arthur Benke, an aquatic ecologist at the University of Alabama in Tuscaloosa, who called the strong language in some of the criticisms "inappropriate".

What drives the critics? Financial or professional ties to the biotech industry don't seem to be the impetus. Such ties do exist — like many people researching biotech crops, some have received research grants from industry or have other interactions with it — but in interviews they say that these are not the major driving force. Rather, many of them feel strongly that transgenic crops are safe and beneficial to the environment and society, and that the image and regulation of these crops has been too harsh. Many of the critics have been studying biotech crops since they were developed commercially in the late 1980s, and some were involved with the first regulatory approvals. They have specific ideas about how the risks of these crops should be scientifically assessed. And they worry that papers that fall short of high standards will give anti-GMO activists ammunition to influence policy, just as the

Protesters can brandish science suggesting that genetically modified crops are harmful.



monarch-butterfly study did. "When bad science is used to justify bad public policies, we all lose," says McHughen, who says he is on a "campaign to make academic scientists a little less politically naive and a bit more careful in their scientific work". Miller adds that "agricultural biotech has been so horrendously, unscientifically regulated and so over-regulated and so inhibited over the past 30 years that to have these pseudo-controversies stirred up unnecessarily does a disservice to everyone and everything".

Ammann points to the example of golden rice, a variety engineered in the late 1990s to contain more vitamin A. Regulations have delayed the rice's development, he says, although more than 250,000 children a year go blind from vitamin-A deficiency. "We have to get emotional," says Ammann. "I can't agree with the cool scientists' perspective — only dealing with the facts. We live in the real world." In 2006, Ammann formed a rebuttal team called ASK-FORCE to challenge reports about biosafety of GM crops. On one online site, Ammann criticizes 20 reports — none of them positive toward biotech crops — that he considers biased or bad science. In July, he was revising a critique of a paper that appeared in *The Lancet* ten years ago. "I'm working nearly day and night on these things," says Ammann.

The emotional and sometimes harsh quality of some of the attacks strikes some scientists as strange and unlike the constructive criticism to which they are accustomed.

P. PAVANI/AFP/GETTY



Gabor Lövei reviewed lab experiments on the impact of biotech crops on predators of crop pests. Pictured, an assassin bug.

P. SHIRK

“inappropriate”. For example, it says the authors didn’t distinguish whether predators and parasites of insects that feed on biotech crops were affected by the toxins in the plants or by the health of their prey. Lövei and his co-authors say they hope to defend their paper in the October issue of *Environmental Entomology*, and will agree that the distinction would have been useful, but that it would not have changed their conclusions.

Shelton says that he and his group wanted to counteract any effect Lövei’s work might have on policy, particularly as that month the European Food Safety Authority was writing up a risk assessment on unintended effects of genetically modified plants. “I could envision a regulator having this Lövei article appear on his desk and saying ‘We’ve got to rethink approval methods,’” says Shelton.

Shelton’s critique was a “way over-reaction”, says an editor at the Entomological Society of America, which publishes *Environmental Entomology*, who asked to remain anonymous. “They seem to have read it with eyes predisposed to

dismiss anything reflecting poorly on GMOs [genetically modified organisms].” Shelton disagrees. “I have been critical of some aspects of genetically engineered plants and microbes in the past when I thought they were warranted, based on scientific data,” he says. “I am also an editor in the Entomological Society of America and felt that our reaction was not a ‘way over-reaction’.” He adds that as an editor he routinely rejects papers that could be considered supportive of GMOs because of their quality. “Poor science can occur on both sides of issues.”

When asked to point to a good paper that reflects negatively on biotech crops, most critics *Nature* spoke to said they couldn’t name any. “I have seen very little substantive data that are negative towards *Bt* crops that can’t be easily overturned,” says Moar. Wayne Parrott at the University of Georgia in Athens, however, says: “There is plenty of biotech-safety research out there that has not come under attack, even when the answers are not what everyone would have liked.”

E.W.

Benke points out that none of the criticisms on the caddisfly paper, for example, called for further study on the insects. “What papers like this do is alert us to possible reasons to look into this more carefully,” he says. “No one mentioned this.” To try to dismiss the research out of hand ignores how science is supposed to work, adds Power — you make a hypothesis, test it, refine it, test it and refine it again. “You keep doing that until you have an answer that is as close as you’re going to get,” she says. “I don’t understand the resistance to that notion.”

Arbiters of the truth

Some scientists say they are galled by the certainty with which some of the critics state their opinion. “Part of what exasperates me is that they have declared themselves to be the experts in this field, and forcefully present themselves as the ultimate arbiters of truth,” says an editor for the Entomological Society of America who asked to remain anonymous. “I personally am in favour of GMOs in general, and think that they are very beneficial for the environment. But I do have problems with the tactics of the large block of scientists who denigrate research by other legitimate scientists in a knee-jerk, partisan, emotional way that is not helpful in advancing knowledge and is outside the ideals of scientific inquiry.”

The critics respond that they are simply pointing out flaws in research, and that this is an important part of the scientific process. “It is neither fair nor accurate to equate pointing out serious deficiencies with experimental design

and data interpretation as ‘denigration,’” Parrott says. “For science to maintain its integrity and move forward, it is critical to assert the right of scientists to question each other’s work.” McHughen says that he doesn’t condone *ad hominem* attacks. “They are invariably unproductive,” he says, and points out these tactics are often used against scientists who don’t oppose GM crops.

Federici says he finds it inappropriate to call the reactions ‘knee-jerk’ ones. “Losey and colleagues, and Rosi-Marshall and colleagues at the time of their studies were newcomers to the field. Most of the people who found their studies flawed and protested had extensive experience with *Bacillus thuringiensis*.” He also points out that the critics varied in how strongly they responded to the Rosi-Marshall paper, saying “I don’t consider writing a letter to the editor a harsh response.”

Ignacio Chapela, a microbial ecologist at the University of California, Berkeley, says that the attacks may be deterring young scientists from pursuing careers in biotech crop research. “I have a very long experience now with young people coming to me to say that they are not going into this field precisely because they are discouraged by what they see,” he says. Chapela faced criticism from pro-GMO scientists after publishing a 2001 paper in *Nature*, in which he reported that native maize varieties in Mexico had been contaminated with transgenic genes¹³. Following the criticism, *Nature* decided that “the evidence available is not sufficient to justify the publication of the original paper”.

At its worst, the behaviour could make for a downward

“When bad science is used to justify bad policies, we all lose.”

— Alan McHughen

spiral of GM research as a whole, says Don Huber, a emeritus professor of plant pathology at Purdue University in West Lafayette, Indiana. "When scientists become afraid to even ask the questions ... that's a serious impediment to our progress," he says. Miller says: "I don't see how criticism of flawed science that verges on misconduct should discourage anybody." Researchers could be invigorated by entering a field with such lively debate. "For some people it might be exciting because you're doing science that is relevant to society," says Power.

Pervasive spread

Rosi-Marshall's caddis-fly paper did find its way into the anti-GMO rhetoric, although on nowhere near the scale that the monarch butterfly paper did. For example, the London-based Institute of Science in Society, a not-for-profit organization involved in the GM debate, on 30 October 2007 posted its summary of the paper, saying that: "calling a halt to planting *Bt* corn next to streams ... would be in keeping with the evidence [the authors] have provided". Greenpeace included the paper in an April 2008 briefing on *Bt* maize, citing it as evidence of environmental risk.

The impact went further than that. On 9 January 2008, three months after Rosi-Marshall's paper was published, France's watchdog on GM foods ruled that one of Monsanto's types of *Bt* maize, known as MON810, may have an impact on wildlife. The evidence it cited included Rosi-Marshall's paper. Two days later, the French government announced a ban on cultivating the maize. "[The paper] got to every agency and non-governmental organization that doesn't like the technology and gave them a flag to wave," says Parrott. Not that he considers the effort wasted: "I have no doubt the impact on policy-makers would have been much worse had it not been countered."

Nearly two years since the paper was published, the critics' comments are still pointed. "It was just an idiotic experiment," Miller said this July. But Rosi-Marshall and her co-authors stand behind their paper. "We believe our study was scientifically sound," they wrote in an e-mail, "although many questions on the topic remain to be answered. The

"Young people are not going into this field because they are discouraged by what they see."

— Ignacio Chapela

repeated, and apparently orchestrated, *ad hominem* and unfounded attacks by a group of genetic engineering proponents has done little to advance our understanding of the potential ecological impacts of transgenic corn."

And Rosi-Marshall's career seems to have survived the furore. In May 2009 she secured tenure at Loyola University Chicago, and in August she moved to the Cary Institute of Ecosystem Studies in Millbrook, New York. There she will study human-dominated ecosystems and will continue to investigate the influence of maize varieties on stream ecosystems. Since the caddis-fly paper, she has co-authored another study on transgenic crops showing that *Bt* maize debris decomposes in streams at a faster rate than conventional maize¹⁴. She says more data produced with the NSF grant are on the way and that the attacks won't deter her from her studies.

"It toughened me up a lot," she says. "I'm not going to be intimidated."

Emily Waltz is a freelance writer based in New York City.

1. Rosi-Marshall, E. J. et al. *Proc. Natl Acad. Sci. USA* **104**, 16204–16208 (2007).
2. Schubert, D. *Nature Biotechnol.* **20**, 969 (2002).
3. Losey, J. E., Rayor, L. S. & Carter, M. E. *Nature* **399**, 214 (1999).
4. Scriber, J. M. *Proc. Natl Acad. Sci. USA* **98**, 12328–12330 (2001).
5. http://pubresreg.org/index.php?option=com_content&task=view&id=64
6. Parrott, W. *Proc. Natl Acad. Sci. USA* **105**, E10 (2008).
7. Beachy, R. N., Federoff, N. V., Goldberg, R. B. & McHughen, A. *Proc. Natl Acad. Sci. USA* **105**, E9 (2008).
8. *Technical Review: Rosi-Marshall et al. 2007. PNAS* **104**: 16204–16208 (Monsanto, 2007); available at http://www.monsanto.com/pdf/products/caddisflies_review_810.pdf
9. Miller, H. I., Morandini, P. & Ammann, K. *Trends Biotechnol.* **26**, 122–125 (2008).
10. Shantharam, S., Sullia, S. B. & Shivakumara Swamy, G. *Curr. Sci.* **95**, 167–168 (2008).
11. Rosi-Marshall, E. J., Tank, J. L., Royer, T. V. & Whiles, M. R. *Proc. Natl Acad. Sci. USA* **105**, E11 (2008).
12. Pokelšek, J. D. et al. Presentation at the North American Benthological Society Annual meeting, Columbia, South Carolina, 2007. Available at www.benthos.org/database/allnabstracts.cfm/db/Columbia2007abstracts/id/370
13. Quist, D. & Chapela, I. H. *Nature* **414**, 541–543 (2001).
14. Griffiths, N. A. et al. *Ecol. Appl.* **19**, 133–142 (2009).
15. Beachy, R. et al. *Nature Biotechnol.* **20**, 1195–1196 (2002).
16. Tabashnik, B. E., Gassmann, A. J., Crowder, D. W. & Carrière, Y. *Nature Biotechnol.* **26**, 199–202 (2008).
17. Moar, W. et al. *Nature Biotechnol.* **26**, 1072–1074 (2008).
18. Lövei, G. L., Andow, D. A. & Arpaia, S. *Environ. Entomol.* **38**, 293–306 (2009).
19. Shelton, A. M. et al. *Transgenic Res.* **18**, 317–322 (2009).





TUNDRA'S BURNING

Lightning and fires on the Arctic tundra seem to be on the rise. **Jane Qiu** meets the researchers learning from the scorched earth in Alaska.

More than 20,000 lightning strikes were recorded on the North Slope of Alaska in 2007. Some struck the vast stretches of lakes; some hit the treeless tundra. And one of them torched into life the largest and longest-lasting tundra fire recorded in the state's history. The blaze, which started near the Anaktuvuk River on 16 July, burned 7,000 hectares a day at its peak, and eventually consumed 100,000 hectares, an area larger than that of New York City. It finally stopped burning in early October, smothered by thick snow.

Two years later, the scars left by the blaze are all too apparent from a helicopter circling over the region. So too is the area's quick recovery. Tussock grass, the predominant vegetation in northern Alaska, sends up vibrant green shoots from scorched meristems. Its white flowers bloom over the deeply blackened soil like a dust of snow, stretching to a hazy horizon. It is surprisingly beautiful.

This is more than a view of nature's swift destruction and renewal; it is also a site of intense research. "The Anaktuvuk River fire is a large natural experiment," says Gaius Shaver, an ecologist at the Marine Biological Laboratory in Woods Hole, Massachusetts, who leads an effort funded by the US National Science Foundation to study the fire's environmental impact. "It provides an unprecedented opportunity

to study how the entire ecosystem responds to major disturbances." Scientists from ten research groups have been flying into the burned area from the nearby Toolik research station to assess how the fire has shifted the carbon balance and affected the hill slopes, valleys, streams and lakes in the region.

Understanding the effect of fires on the Arctic tundra may become more important as the climate gets warmer. Tundra fires used to be rare events, but higher temperatures and a more arid climate seem to be changing that. According to the US Bureau of Land Management in Washington DC, the frequency of lightning on the North Slope has increased tenfold in the past decade. And many researchers fear that the increased lightning may increase the fire risk. Of the 26 recorded fires on the North Slope since 1950, dose to one-third has taken place in the past three years — and the region burned in the Anaktuvuk River fire alone constitutes more than half of the total burned area.

Just as climate change may fuel fires, fires may accelerate climate change. These vast areas of tundra store about 14% of the world's soil carbon at the surface alone. Fires could release a large amount of that, either directly through combustion or indirectly by modifying the tundra ecosystem. Many of the studies in Alaska are designed to look at how this could play out.

"Tundra fires could significantly impact the carbon balance of the Arctic and exacerbate global warming," Shaver says.

The first pulse

One of the first researchers to arrive in the area after the fire was Cody Johnson, an aquatic ecologist at Utah State University in Logan, along with a small team. The razed landscape had been buried under snow all winter. And it still wasn't exactly camping season when, in May 2008, they flew out to Dimple Lake in northern Alaska, at 69° latitude and 200 miles north of the Arctic circle. The crew landed in skis, dragged their heavy equipment to the lake shore, set up camp and waited in sub-zero temperatures for the first snow melt.

After two weeks, Johnson could see the land for the first time. "The surface was covered in a thick layer of ash," he says. "There was a pungent smell of the burn in the air." Underneath the ash they found a blackened land with islands of yellow tussock trunks sticking out. These many-layered skirts of grass are so thick and packed so tightly that the core never dried out, saving them from the fire.

The main purpose of Johnson's camping trip was to study the movement of burned materials, especially ash, nutrients and sediments, into streams and lakes, and its effect on the ecosystem's carbon cycle. After the plants that hold

R. GLENN

the soil together have burned, these nutrients and sediments can be readily flushed out. A boost in nutrient levels could stimulate the rate of primary production — the production of organic compounds from the fixing of carbon dioxide during photosynthesis. By contrast, additional sediments could partly offset this effect by smothering algae, the key player in carbon fixation. Working out the net outcome of these and other interconnecting processes on carbon is no mean feat.

George Kling, a lake ecologist at the University of Michigan in Ann Arbor, says that the lakes in the Arctic, already supersaturated with carbon, are a net source of carbon dioxide and methane — they release more of these gases than they can store. He and others have shown that when carbon moves from land into lakes, some 20% of the Arctic carbon sink is re-released¹. “You can think of the aquatic system as a release valve, which leaks carbon from land to water and then back to the atmosphere,” says Kling, who is collaborating with Johnson. “Major disturbances such as the Anaktuvuk River fire could turn that trickle into a large flood and cause much greater movement of carbon.”

The researchers predicted that the biggest inflow of nutrients after the fire would be from the snow run-off. “We wanted to capture that first pulse,” says Johnson. Every day they sampled what went into the streams and how much of those materials ended up in rivers and lakes. Their initial results, which have not yet been published, show that streams in the burned areas dumped up to five times more sediment into Dimple Lake than those in unburned ones, but carried only twice as much carbon. Puzzled by the discrepancy, Kling and his colleagues looked at the quality of the water that leached out from the soil and found that microbes performed decomposition seven times faster in water from the burned area than in unaffected sites. “It seems that carbon is released into the atmosphere through microbial decomposition before it reaches the streams,” says Kling.

Johnson and his team weren't camping alone for long. In June 2008, Michelle Mack, an ecologist at the University of Florida in Gainesville, flew to the burned area with a handful of colleagues to reconstruct how much carbon and nitrogen were there before the fire, and to calculate how much carbon was released into the atmosphere during and after it. By using tussock meristems as

measuring sticks, they estimated the depth of the soil before the fire. They also gauged the carbon concentration in the soil from unburned sites at the same climate and elevation as the burned area.

The blaze generated a big carbon cloud. Mack found that the fire emitted 1.8 million tonnes of carbon dioxide into the atmosphere, equivalent to about 0.03% of the annual global carbon emission due to human activity. “That amount was certainly large enough to change the carbon balance of the entire North Slope for 2007,” Shaver says.

Burning up

The burned area continues to pump carbon into the atmosphere, says Adrian Rocha, an ecologist working in Shaver's group. To measure the net carbon flux, Rocha and his colleagues set up a meteorological tower in a severely burned area, a moderately burned one and an unburned control site. They found that, within a radius of one kilometre around the tower, the severely burned site released 80 grams of carbon per square metre during the summer months, whereas the unburned tundra was absorbing 50 grams of carbon per square metre. They suspect that, in the burned area, the rate of carbon fixation by photosynthesis dropped because so many plants were destroyed, and that was outweighed by the rate of carbon being released by microbial decomposition in warmer soil. Now Rocha is extrapolating the results to a larger area. He is using a satellite instrument called the Moderate

Resolution Imaging Spectroradiometer (MODIS) to measure how green the land surface is, and is building a model to correlate the greenness to the type of tundra vegetation².

Once he has a good model, he will be able to estimate the carbon flux of the entire burned area for different periods after the fire.

Mack found that 60% of the initial carbon emission was from the burning of organic plant and animal material in the soil — which is usually pretty wet on the tundra and doesn't make an efficient fire fuel. The rest came from mosses, lichens and other plant materials on the tundra surface. The organic

matter in soil insulates the permafrost — the permanently frozen ground — underneath, and any reduction in the soil depth could mean that the underlying permafrost warms and thaws, with big repercussions for the ecosystem. At the end of June 2009, the permafrost had thawed to a depth of 40 centimetres in severely burned areas, compared with 18 centimetres in unburned sites. Permafrost thaw can be exacerbated by the fire-blackened surface, something else that Rocha and his colleagues are measuring.

Researchers have been trying to understand the impact of permafrost thaw on carbon exchange for some time. A study by Edward Schuur, an ecologist at the University of Florida, and his colleagues, showed that there was net carbon loss from a tundra landscape that had undergone permafrost thaw for more than several decades³. The world's permafrost contains

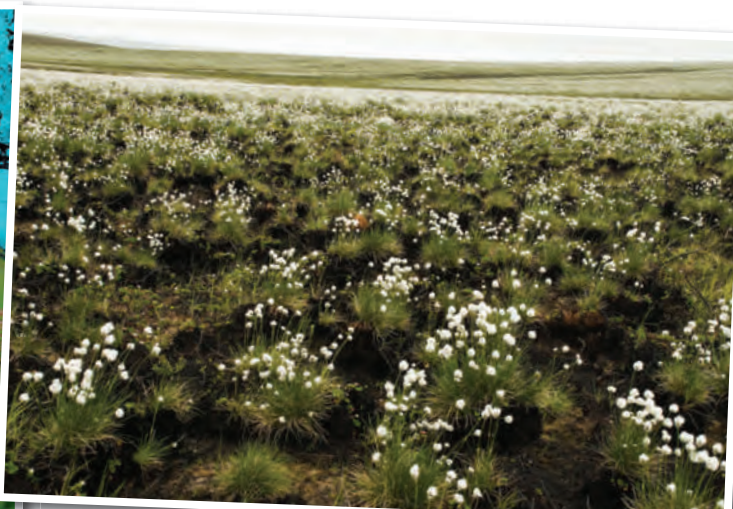
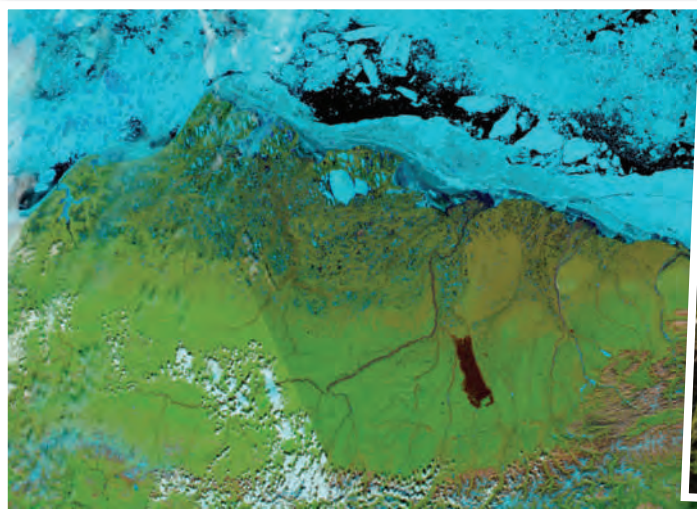
twice as much carbon as its atmosphere; it would be bad news if even a fraction of that were released⁴.

The accelerated melting of permafrost caused by the Anaktuvuk River fire has carved more visible changes out of the landscape. The newly softened soil can collapse or slide away, creating a landscape that is pock-marked with hollows known as thermokarsts. Johnson noticed thermokarsts forming in autumn 2008, and they had grown much bigger when he went back this summer. He watched one, on the shore of Horn Lake, double in width to 200 metres in two days. In one valley, which has



A fire started by lightning blackened vast tundra regions near the Anaktuvuk River.

B. BOWDEN



Studies are exploring the 100,000 hectares burned in the Anaktuvuk River fire (brown rectangle, left); cotton-grass flowers were one of the first signs of recovery.

NASA-MODIS

now been dubbed 'the valley of thermokarsts', there are about a dozen on each slope. "It's a muddy holocaust," says Kling.

Many fear that major disturbances to the Arctic, such as fires and the resulting thermokarsts, could change the vegetation on the tundra, so they are monitoring it closely as it rebounds. Tussocks are amazingly resilient and even the most severely burned ones sprouted as soon as the spring kicked in last year. Sphagnum moss and fruticose lichens, the 'down jacket' of the tundra that insulates the soil from incoming heat, have shown no sign of recovery. In their place are patches of copper-wire moss, with its red stems and golden, hanging capsules, and liverwort, holding up its peculiar umbrella-shaped reproductive organs.

Spread of the shrubbery

The researchers worry that exposed mineral soil and more nutrient availability after fires will favour the expansion of species such as shrubs. This, too, could alter the net carbon exchange: studies of experimental plots show that more nutrients are associated with more shrubs and a net loss of soil carbon. But there are issues beyond that. Shrubs are drier and woodier so burn better than most other tundra vegetation — more shrubs growing after this fire could therefore push up the risk of future ones. "Shrub expansion could be a positive feedback to climate change and fire frequency and severity," says Charles Racine, an ecologist who has been studying fires in Alaska since 1977, and who has retired from the Cold Regions Research and Engineering Laboratory in Hanover, New Hampshire.

In collaboration with Randi Jandt of the Alaska Fire Service at Fort Wainwright, Racine and his colleagues have documented the vegetation recovery after the 1977 tundra

fire on the Seward Peninsula in northwestern Alaska, which left less severe damage than the Anaktuvuk River fire even though a similarly large area was burned⁵. Thirty-two years after the fire, there is still no sight of the sphagnum moss or fruticose lichens, which before the fire covered 20% and 7% of the surface, respectively. By contrast, the area covered by willow, a deciduous shrub, rose from 5% in 1973 to 25% in 2001 and to about 40% this year. Philip Higuera, a palaeoecologist at the University of Idaho in Moscow, Idaho, and his colleagues looked further back in history. By analysing the pollen and charcoal contents in sediments from lakes in northern-central Alaska that are 15,000 years old, they found that an increase in fire frequency coincided with the vegetation transition from herbs to shrubs around 14,000 years ago⁶.

Whatever fuelled the blazes of the past,



Soil softened by fire can collapse, creating thermokarsts that punctuate the landscape.

researchers are keen to find out what kept the Anaktuvuk River fire burning, and how others could be predicted in the future. The summer of 2007 saw the record high temperature and record low precipitation north of the Brooks mountain range, resulting in extremely dry soil⁷. Yet there were two other fires near the Anaktuvuk River that went out rapidly.

Nancy French, an expert on remote sensing at the Michigan Tech Research Institute in Ann Arbor, and her collaborators, including Jandt and Racine, are hoping to get a grant from NASA to map past tundra fires across Alaska and Canada using satellite data. "There is an urgent need to have a comprehensive record of tundra fires, so we could be in a better position to predict tundra fire regimes in the future," she says. The researchers plan to correlate tundra fires with weather conditions, and to construct computer models to calculate carbon emission during a fire, predict vegetation recovery and test how climate change could affect fires in the tundra.

For the researchers at Anaktuvuk River, meanwhile, the tundra is changing almost daily. On the northeastern bank of Dimple Lake, a massive patch of soil has slumped into the water, muddying the turquoise with grey. Farther up the shore, the cotton grass ripples across the hill in a breeze. This is the scene after just one lightning strike sparked a fire. And researchers can be confident that lightning will strike again. ■

Jane Qiu, a recipient of the 2009 Marine Biological Laboratory Logan Science Journalism Fellowship, writes for Nature from Beijing.

1. Kling, G. W., Kipphut, G. W. & Miller, M. C. *Science* **251**, 298–301 (1991).
2. Rocha, A. V. & Shaver, G. R. *Agric. For. Meteorol.* **149**, 1560–1563 (2009).
3. Schuur, E. A. G. et al. *Nature* **459**, 556–559 (2009).
4. Walker, G. *Nature* **446**, 718–721 (2007).
5. Racine, C., Jandt, R., Meyers, C. & Dennis, J. *Arctic Antarct. Alp. Res.* **36**, 1–10 (2004).
6. Higuera, P. E. et al. *PLoS ONE* **3**, e0001744 (2008).
7. Jones, B. M. et al. *Arctic Antarct. Alp. Res.* **41**, 309–316 (2009).

J. QIU

B. BOWDEN

CORRESPONDENCE

Conservation: a small price for long-term economic well-being

SIR — Public funds should not just be channelled into boosting immediate economic recovery — we must also be prepared to pay a higher ‘insurance bill’ to safeguard the ecological long-term basis of our economic and social well-being. Increased spending on conservation could help fund a worldwide core network of protected areas of biodiversity.

At the beginning of this year, worldwide economic stimulus plans totalled US\$2.8 trillion, with some 15.6% allocated to ‘green’ purposes such as climate protection (<http://tinyurl.com/ddrj69>). In April, the G20 group of nations announced plans to spend US\$5 trillion by the end of 2010, to stimulate the floundering global economy and “accelerate the transition to a green economy”. None of these funds is designated to the conservation of biological diversity — a fundamental resource for our economy and society.

The multilateral Global Environmental Facility (an independent financial organization uniting 178 member governments, in partnership with international institutions, non-governmental organizations and the private sector) has spent US\$7.6 billion on biodiversity conservation since 1991. In addition, we estimate that some \$7.6 billion has been contributed to this purpose each year by national governments.

We estimate that US\$53 billion is needed annually to maintain a global core network of protected areas (our update of figures from A. Balmford *et al.* *Science* **297**, 950–953; 2002). This amount includes the income opportunities lost for the rural population of poor countries: mainly income that is, or could be, generated from tropical and subtropical agriculture.

It has been known for some

time that the public in wealthy countries would be prepared to contribute additional funding to promote biodiversity conservation in the tropics (see, for example, B. Horton *et al.* *Environ. Conserv.* **30**, 139–146; 2003). We believe that these contributions would today amount to US\$42–140 per household per year. Extrapolating this to each of about 452 million households of the member countries of the Organisation for Economic Co-operation and Development, we believe that roughly US\$43 billion could be generated annually. This sum would cover most of the global core network of sites earmarked for protection.

A dramatic increase in conservation spending is no fast fix for the economy. But conserving biological diversity contributes crucially to maintaining the resilience and adaptive capacity of the biosphere. In addition, it directly provides a multitude of critical ecological services (see, for example, <http://tinyurl.com/oe799>).

Boris M. Hillmann, Jan Barkmann
Georg-August-Universität Göttingen,
Department of Agricultural Economics
and Rural Development,
Platz der Göttinger Sieben 5,
37073 Göttingen, Germany
e-mail: bhillma@gwdg.de

Conservation: the world's religions can help

SIR — The world's religions are emerging as a surprising driver of support for conservation of biological diversity.

The International Interfaith Investment Group, for example, which is collectively worth more than US\$7 trillion, is encouraging religious organizations to change their current investment policies in favour of those that support conservation (www.3ignet.org).

In addition, lands owned by these organizations can

contribute to the conservation of biodiversity because of their protected status. More than 7% of Earth's land surface is owned by religious institutions, and a further 8% has sacred links (www.arcworld.org). Given that most countries will never be able to designate more than 15% of their land as protected areas (S. Chape *et al.* *Phil. Trans. R. Soc. B* **360**, 443–455; 2005), territory with religious and sacred affiliations contributes substantially to maintaining biodiversity.

It should also be possible to raise funding for conservation by appealing to donors who have religious faith. For example, the wealthy countries of the G20 group that have large religious populations might step in and help.

The focus of initiatives in the past has been on paying for ecosystem services, which are considered ‘natural capital’ (R. Costanza *et al.* *Nature* **387**, 253–260; 1997), but an appeal to support native communities on religious grounds might prove more persuasive in a difficult economic climate.

Of the 125 countries that are represented in the Conservation International list of biodiversity hotspots (www.biodiversityhotspots.org), most have a low per-capita gross domestic product (GDP) and a strong religious base (<http://tinyurl.com/2b2kg9>). Collectively, these countries are home to more than 4 billion people affiliated with one of 11 mainstream faiths; more than half of them have a total population of 3 billion and a per-capita GDP of less than US\$5,000.

Religious sympathy has the potential to make a major contribution towards biodiversity conservation. This contribution could be extremely valuable in the approach to the 2010 target of the Convention on Biological Diversity.
Shonil Bhagwat School of Geography and the Environment, University of Oxford, South Parks Road, Oxford OX1 3QY, UK

e-mail: shonil.bhagwat@ouce.ox.ac.uk
Martin Palmer Alliance of Religions and Conservation, The House, Kelston Park, Bath BA1 9AE, UK

Defining numbers in terms of their divisors

SIR — In her informative Book Review of Yoko Ogawa's *The Housekeeper and the Professor* (*Nature* **460**, 461–462; 2009), Jennifer Rohn nicely illustrates how one false premise will lead to all kinds of misunderstanding. She states that primes resist “division by any number other than zero and one”. She means, of course, that a prime number resists division by any number other than itself and one. Every integer can count on two divisors: itself and one. All resist division by zero.

If we do not exclude unity and the number itself when summing up divisors, the ‘abundant’ numbers, whose divisors’ sum is greater than themselves (for example, the number 18 referred to in the review), really would be abundant — and would include all numbers higher than one. Meanwhile, ‘deficient’ numbers such as the number 14, for which the divisors’ sum is less, would be non-existent.

D. Speijer Academic Medical Centre, University of Amsterdam, Medical Biochemistry, Meibergdreef 15, 1105 AZ Amsterdam, the Netherlands
e-mail: d.speijer@amc.uva.nl

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than about 300 words, and ideally shorter, with no more than three references and three authors (for details, see <http://tinyurl.com/373jsv>). Published contributions are edited. Science publishing issues of interest to authors are regularly featured at Nautilus (<http://blogs.nature.com/nautilus>), where we welcome comments and debate.

OPINION

Pakistan's reform experiment

In 2002, Pakistan began an ambitious overhaul of its higher-education system. The successes and failures of the experience hold lessons for other countries, say **Athar Osama** and co-authors.

After decades of neglect, in 2002 Pakistan set out to dramatically reform its higher-education system. The reforms were designed to reverse years of chronic underfunding, to invest in the academic workforce and to revitalize a moribund research enterprise. This ambitious agenda generated immense public interest and controversy. Although it is too early to judge the outcome of the experiment, it is already clear that some initiatives were more successful than others.

Highs include more research papers, more PhDs and greater visibility for Pakistani research. The lows include an absence of external peer evaluation and of rigorous impact metrics. At times the speed and scale of reform outpaced the ability of Pakistan's universities to adapt. And the top-down nature of the revamp also led to distress among faculty members. An important lesson for would-be reformers is that greater participation and openness may increase credibility and sustain support for reforms.

Higher education in Pakistan has been a story of neglect for much of the country's 62 years. Outside a few pockets of excellence, many of the

institutions have been marked by mediocrity and a lack of motivation. Rather than contributing to the creation of new knowledge, they have been institutions of rote learning and feeder schools for foreign universities.

After rising to power in 1999, the military government of General Pervez Musharraf undertook a series of reforms. This coincided with broader international support from development and donor communities for investment in higher education. The public-private Task Force for Higher Education was set up in 2001 to review Pakistan's higher-education situation and recommend options for improvement.

The task force made some startling discoveries¹. In Pakistan, only 2.6% of 17–23 year olds were enrolled in an institution of higher education compared with 6.2% in India, 12.7% in Iran and 68% in South Korea. Of those few, only about 30% went to universities with the rest attending two-year colleges. The total public funding of 41 public-sector universities was a meagre Rs3.9 billion per annum (US\$55 million in 2001 dollars). This amounted to a mere 0.1% of the country's gross domestic product².

SUMMARY

- Big investment in higher education in Pakistan over the past 7 years
- Papers and PhDs up, engineering and maths show biggest improvement
- Education commission needs an independent review of its policies

The national annual allocation for university research was only Rs40 million¹.

Chronic underfunding of higher education was just one of the challenges identified. Other concerns were a lack of political will for meaningful reform, a lack of appreciation for the role education can play in development, ineffective governance systems, political interference in university administration, weak institutional leadership and, at the university level, a lack of performance culture and accountability.

March of progress

The task force made a series of recommendations, including calling for a huge jump in public funding for universities from Rs3.9 billion in 2001–02 to Rs21.7 billion in 2005–06 (ref. 2). An action plan was taken forwards by the Higher Education Commission (HEC), which replaced the University Grants Commission.

Starting in 2003, the HEC launched programmes in human-resources development, research and physical infrastructure, and reforms of curriculum, governance and pedagogy (see 'Pakistan's key initiatives — progress so far'). Human resources took the lion's share of investment, and often received the strongest criticisms. Inevitably, some programmes were better designed and executed than others.

For example, a foreign PhD fellowship programme has sponsored more than 2,000 scholars to study abroad. To date, the host countries seem to be happy with the quality of these students, although the programme's impact will depend on Pakistan's ability to attract back and reabsorb the scholars.

By contrast, the domestic PhD fellowship programme has had a bumpy start. Here the goal was to create 5,000 new PhDs at local universities over 5 years — from a baseline of a few hundred PhDs in previous years. In this instance, the HEC's critics argue that undue emphasis has been placed on quantity rather than quality. Two factors are at the root of the

PAKISTAN'S KEY INITIATIVES — PROGRESS SO FAR

| Human resources (faculty) development | |
|---|--|
| Foreign PhD fellowships | 2,825 scholarships awarded for study abroad in 20 host countries |
| Foreign Faculty Hiring Programme (FFHP) | Internationally comparable salaries offered to attract foreign faculty members — often Pakistanis working abroad — to take temporary appointments at a public university. 289 foreign professors have joined |
| Tenure Track System (TTS) | Internationally comparable salaries offered to faculty members to enter into a performance contract under the tenure track stream; adoption is marginal |
| Domestic PhD fellowships | To date, 3,516 PhD scholarships awarded — 179 PhDs completed |
| Research and physical infrastructure | |
| National Research Programme for Universities (NRPU) | A peer-reviewed research-grant programme for 3-year projects; 644 projects approved (approval rate: 47%) between 2002–08 |
| Pakistan Education and Research Network (PERN) | High-speed Internet connectivity to 95 universities to enable research collaboration and video-conferencing |
| National Digital Library (NDL) | Free electronic access to 45,000 e-books and 23,000 full-text journals across more than 250 institutions |
| Curriculum, governance and learning innovation | |
| Curriculum reform | Curriculum of 97 disciplines updated; graduate courses increased to 4 years |
| Quality assurance | Quality Enhancement Cells established in 30 universities |
| Governance | Criteria for appointment of university vice-chancellors established; governance and leadership training for university leaders |
| Learning innovation | Programmes to train faculty members in innovative teaching methods |

Source: HEC data (especially ref. 5)

criticism — strong financial incentives for faculty members for each student that they advise, and low entry criteria for students. The HEC's attempts to address concerns have been opposed by some faculty members, perhaps because they were benefitting from the incentives. The result could be too many ill-prepared students gaining doctorates, which could have long-term consequences for the market and for the reputation of Pakistani PhDs.

Arguably, in this and in a few other cases, the HEC adopted a much more aggressive approach to reform than it — or Pakistan's university system — could manage. In some instances, the HEC has been slow to realize the unintended consequences of its programmes. Excessive centralization of the reform effort — which the HEC justified as necessary to keep up momentum — also undermined university leadership and academic freedom.

The commission's attempts to win over the resistance has had mixed results. For instance, by tying faculty performance to financial compensation it created deep divisions between winners and losers in a two-tier salary structure. And, by sending young scholars abroad by the thousands — to be reabsorbed on their return — it created a fresh base of support for its policies. However, support is far from universal — older and relatively independent universities have generally been more critical of HEC encroachment on institutional and academic freedom, and some only selectively participated in HEC programmes³.

Despite this resistance, the HEC seems to have changed the culture of Pakistani academia considerably over the past 5 years³. The HEC claims to have caused a 400% increase in the number of papers published in international journals by Pakistani universities. It also takes credit for the appearance of three Pakistani universities among a popular top-600 chart of world universities, the ranking of Pakistan as a 'rising star' in five fields of science and engineering⁴ and external endorsements by evaluation teams from the British Council, the World Bank and USAID.

It is more difficult to estimate the influence of the HEC programmes on research quality. Metrics are lacking and impact takes longer to become visible. Comparing the pre- and post-reform publishing performance of Pakistani authors, we found that in most fields, Pakistan's impact has improved relative to the world average (see 'Pakistan's academic publishing performance'). Engineering and mathematics seem to have seen the biggest improvement. In mathematics, for example, an average paper by a Pakistani author is cited around 20% more

PAKISTAN'S ACADEMIC PUBLISHING PERFORMANCE

| | Pre-reform (1998–2002) | | | Post-reform (2003–07) | | |
|--------------------------|------------------------|-------------|------------------|-----------------------|-------------|-----------------|
| | Total papers | % of world | Relative impact* | Total papers | % of world | Relative impact |
| Biology and biochemistry | 134 | 0.05 | –77.5 | 282 | 0.09 | –60.9 |
| Chemistry | 725 | 0.14 | –75.9 | 1,582 | 0.26 | –71.3 |
| Computer science | 9 | 0.02 | –80.0 | 50 | 0.09 | –69.3 |
| Economics and business | 16 | 0.03 | –81.3 | 17 | 0.03 | –88.5 |
| Education | 6 | 0.05 | –66.0 | 7 | 0.05 | –64.2 |
| Engineering | 165 | 0.06 | –55.9 | 518 | 0.14 | –10.1 |
| Mathematics | 27 | 0.04 | –70.2 | 66 | 0.08 | 20.8 |
| Physics | 409 | 0.09 | –58.9 | 729 | 0.14 | –44.9 |
| Plant and animal science | 474 | 0.21 | –68.8 | 863 | 0.34 | –67.6 |
| Social sciences, general | 67 | 0.05 | –12.9 | 117 | 0.07 | –27.9 |

Source: Science Watch/Thomson Reuters. *Percentage a paper is cited more or less than an average paper in the discipline.

than the worldwide average for the discipline.

The strongest criticism of the reforms is that by vesting most powers within one body, the HEC became the initiator, implementer and evaluator, making accountability problematic or impossible. This created opposition from those who might have agreed with the reforms but were opposed to the implementation. Greater transparency and accountability would have diverted some of this criticism.

More consultation and external oversight would have reduced the momentum for reform, but, in some cases, that may have been a good thing. In our view, reform should be evenly paced — even slowed down — to avoid any real or perceived compromise on quality.

Rapid results

Atta-ur-Rahman, the former chairman of the HEC, is unapologetic for being in a hurry to bring Pakistan's academic establishment into the twenty-first century. To him, a more patient approach to building research capacity would simply not get the job done. Under his command, the HEC pursued radical overhaul and chose to focus resources on a few areas, such as natural sciences and engineering, where it could quickly demonstrate results. Although the jury is still out on the wisdom of this approach, it did create the necessary momentum for the reform agenda.

In the current economic environment, with a change in national government, and fresh questions being asked about the HEC's budget and programmes, there is great need for an independent peer review of the HEC's performance. Ideally, this review would be undertaken by a committee of national and international experts in policy and academic governance. This

could position the HEC for a next phase of the reform effort that builds on its more successful programmes and improves on the less successful ones. Any reform agency must be willing to continually learn and adapt. We recommend that the HEC devotes resources to setting up data-collection systems and processes that would allow it to learn from its past experiences.

The HEC has, over the past few years, made considerable progress. Its success, however, must not be measured by the number of grants made or PhDs awarded. Rather it should be judged on whether it is creating a culture of research — one driven not by financial incentives, but by a genuine desire to create new knowledge and to enable the broader society to reap the benefits. While that remains to be seen, Pakistan's experience has useful lessons for other countries. ■

Athar Osama is a visiting fellow and **Adil Najam** is the director of the Pardee Center for the Study of Long Range Future at Boston University, Boston, Massachusetts 02215, USA. **Shamsh Kassim-Lakha** is former minister of education, science and technology, Pakistan. **Syed Zulfikar Gilani** is at the Institute of Education and Research, University of Peshawar, Peshawar, Pakistan. **Christopher King** is editor of ScienceWatch.com, Thomson Reuters, Philadelphia, Pennsylvania 19104, USA. e-mail: athar@atharosama.com

1. *Higher Education in Developing Countries: Peril and Promise* Task Force on Improvement of Higher Education in Pakistan (World Bank and UNESCO, 2002).
2. *Pakistan: An Assessment of the Medium-Term Development Framework* Report No. 37247 (World Bank, 2006).
3. *Evaluation of USAID Higher Education Portfolio* Task Order No. REE-I-07-05-00053-00 (USAID, 2008).
4. *Rising Star Rankings* Science Watch (Thomson Reuters, 2008); available at <http://sciencewatch.com/dr/rs/08sep-rs>
5. *Higher Education Commission Report 2002-2008* (Higher Education Commission, 2008).

See Editorial, page 11.

"The commission should be judged on whether it is creating a culture of research."

BOOKS & ARTS



A member of a Sunni extremist group during a protest in Lahore, Pakistan, last month.

A. ALI/AP/GETTY

Like minds can be small minds

An adviser to US President Barack Obama argues that people's tendency to seek out those with similar views can entrench extreme opinions. But many other forces can fuel outlandish beliefs, says **Herbert Gintis**.

Going to Extremes: How Like Minds Unite and Divide

by Cass R. Sunstein

Oxford University Press: 2009. 208 pp.
\$21.95, £12.99

Law professor Cass Sunstein creatively combines academic scholarship, popular writing and public service. As an adviser to US President Barack Obama, he has recently moved from his position at Harvard Law School to head the Office of Information and Regulatory Affairs. His latest book, *Going to Extremes*, examines how social segregation causes people's views to become more exaggerated.

Sunstein's specialism is behavioural decision theory, a field that has demonstrated that beliefs are not formed using rational deliberation alone. In this book, he gathers evidence to show that when like-minded people interact, their views are reinforced and become more extreme. For instance, in one experiment, subjects who identified themselves as 'liberal' or 'conservative' became even more so after being separated into two groups of like-minded individuals to discuss controversial issues such as global warming, abortion and gay marriage.

Sunstein concludes that people tend to seek others with similar ideas, and their interactions give rise to 'group polarization'. His key example is religious terrorism, as perpetrated by young men who spend time together, pray together

and read the same materials. In reinforcing the legitimacy of one another's complaints, they generate an unwarranted justification for their violent intentions and a groundless optimism in their ability to succeed.

Going to Extremes is also a cautionary tale for Obama. His predecessor, George W. Bush, has been widely criticized for failing to supply sufficient troops and equipment to the rebuilding initiative in Iraq following the fall of Saddam Hussein. Bush's failure to act has often been attributed by the Democratic opposition to the similar mindset of his team of advisers, who were hand-picked to work beside defence secretary Donald Rumsfeld. Very soon after Rumsfeld was replaced by Robert Gates in December 2006, Bush's strategy changed. After much consultation, in January 2007 Bush authorized a dramatic increase in troop levels in Iraq.

Sunstein argues that a diversity of advisory opinion is essential to good leadership, to avoid executive policy becoming entrenched irrespective of unfolding events. As a small group of people all committed to one viewpoint, Bush's advisers moved to a position that was more extreme, and less tenable, than any single adviser would have held in isolation.

However, not all collective false beliefs are the product of self-segregation. 'Belief contagion' can lead large numbers of people to accept outlandish notions for which there are no credible evidence. For example, in the

mid-twentieth century, psychoanalyst Bruno Bettelheim asserted, on the basis of anecdotal evidence, that autism was caused by "refrigerator mothers" — women who treated their children coldly. Widely endorsed at the time but now regarded as erroneous, this runaway belief heralded years of victimization of the mothers of children with autism and misdirected therapeutic efforts.

A more recent example of belief contagion from the late-twentieth century is the preoccupation with 'recovered memories', which are susceptible to suggestion and hence can be easily spread. There have been many cases of teachers, for example, being prosecuted for alleged sexual abuse on the basis of community accusations, backed up with rehearsed child testimony and irregular court proceedings that were shaped by false collective beliefs.

Prejudice is another potent generator of false beliefs. Anti-black sentiments were fuelled in the southern United States after the civil war by self-serving politicians and tabloid newspapers. Under Adolf Hitler, Nazi politicians spread anti-Semitic stories for political reasons, but prejudice rendered these stories plausible.

On reading *Going to Extremes*, one might expect that people are becoming more polarized. But that is not the case. American voters are now more likely to declare themselves Independents, rather than either Democrats or Republicans, than they were a century ago. And in all but four sessions of the US Congress

between 1980 and 2007, the president and the congressional majority have come from different parties. We may feel more polarized today because, especially through the media, we are in daily contact with those with whom we disagree.

Nor should terrorism be thought of as simply the product of group polarization. Terrorism

has long been an effective tactic for non-state actors with political goals, and is supported and funded by large sections of the societies in which they live.

Going to Extremes is a fine book full of insightful evidence and intelligent commentary on modern political life. Sunstein's vision of emancipatory political discourse is salutary,

and the world would probably be a better place if we followed it. But this is only one aspect of how we form beliefs. ■

Herbert Gintis is professor of economics at the Central European University, 1051 Budapest, Hungary, and external professor at the Santa Fe Institute, Santa Fe, New Mexico 87501, USA. e-mail: hgintis@comcast.net

Society need not be selfish

The Age of Empathy: Nature's Lessons for a Kinder Society

by Frans de Waal

Random House: 2009. 320 pp. \$25.99

In the 1950s and 1960s, a major topic of research in animal behaviour was aggression. Konrad Lorenz's popular 1966 book, *On Aggression*, argued that it has an important role in the social life of many animal species, including humans. The idea struck a chord with the public — perhaps because the memory of the Second World War was still fresh and the emerging cold war showed that even allies could turn on one another. Lorenz also noted that many animal species had “aggression-inhibiting mechanisms” to keep things from getting out of hand. But he thought that humans were especially under-equipped with these.

In the late 1970s, ethologist Frans de Waal began to study the aggression-inhibiting mechanisms of various primate species, which led to the discovery that primates often reconcile

after fighting. Combatants will seek out one another and display some kind of peacemaking behaviour that often involves touching, as de Waal summarizes in his 1989 book *Peacemaking Among Primates* (Harvard University Press). The phenomenon has turned out to be robust and widespread, even among some non-primate mammals. De Waal has also studied reciprocity in primates: food sharing, support in fights and grooming seem to be exchanged in kind in many species.

Over the past 15 years or so, de Waal has written a series of popular books with the main theme that humans are more similar to other animals in their attitudes and behaviour than most people realize. In particular, he has set out to combat the idea that evolution selects exclusively for individualistically selfish behaviour in animals. He argues that evolution has also selected for conciliation, cooperation and even empathy in many species, most obviously in those who live in complex social groups, which includes most mammals and almost all primates. To the chagrin of many scientists, de Waal's books often weave together seamlessly

systematic research with informal anecdotes of animal behaviour.

In his latest volume, *The Age of Empathy*, de Waal continues these same themes but with more focus on the implications for how we should conduct ourselves and construct our societies. His evidence is drawn from non-human primates, dolphins, elephants and various domesticated animals, including his own pet cat. He also scrutinizes some politicians and popular figures from the point of view of animal behaviour. A repeated foil throughout is Gordon Gekko, from the 1987 movie *Wall Street*, who reiterates in various forms the basic credo that “greed is good”.

As a European living in the United States, de Waal compares a cooperative, collectivist and caring society with one that is more focused on individuals' rights and responsibilities — confessing his own sentiments to lie “somewhere in the middle of the Atlantic”. His main political message is that we should not continue to harp on about evolution justifying only the selfish side of human nature, although of course that exists. He urges that we must also capitalize on the empathetic and cooperative attitudes that evolution has equipped us with, writing: “A society that ignores these tendencies can't be optimal”.

De Waal's latest book does not address fundamental questions about the evolution of cooperation, empathy or altruism, nor how humans evolved exaggerated forms of these to become fully moral agents. But for those who still equate the terms ‘beastly’ and ‘ape-like’ with fundamentally selfish and aggressive behaviour, and who believe that evolution is always “red in tooth and claw”, *The Age of Empathy* is an excellent antidote. Whether this book, or indeed any study of the natural world, also provides concrete lessons for human society is another question. ■

Michael Tomasello is professor of developmental and comparative psychology at the Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. His forthcoming book is *Why We Cooperate*.



Dolphins' ability to cooperate, argues Frans de Waal, is evidence that evolution favours altruism.

Pop artist displays primitive instincts

A retrospective of Todd Schorr's huge oil-painted comic-book visions features his garish image of a hunter-gatherer. Is it a deliberate allegory of consumer culture, asks **Martin Kemp**?

Sci-fi comics have a look of their own. Their style has changed over the years in response to new print technologies, but they display recurrent visual characteristics that are easy to list. Alien creatures have monstrous heads, leering eyes, savage claws, snaking tentacles and scaly skins. Mechanized organisms do battle with anthropomorphized machines. The paraphernalia and costumes are oddly Gothick; galactic spaces plunge and swirl; and phosphorescent lights compete with devouring darkness. All of this is rendered in lurid colours with exaggerated outlines that shout of violent action and rude excitement.

There seems to be little here to align with 'high art'. However, in the mid-1950s, American and British artists in the pop art movement, such as Andy Warhol and Richard Hamilton, imported imagery and techniques from the commercial world and from 'lowbrow' sources including comics and cartoons. Lowbrow art became a movement in itself, centred on Los Angeles in California from the 1970s onwards, adopting the cartoon style as its major characteristic.

No one has gone to more elaborate lengths than the painter Todd Schorr to transfer lowbrow modes of representation to the 'high' technique of oil painting on canvas — often on a very large scale. Schorr, who is based in Los Angeles, has a background that is perfect for this end.

Starting out as a drummer in rock bands, Schorr was captivated by the culture of psychedelia before training as an illustrator at the Philadelphia College of Art in Pennsylvania. His career flourished: he produced album covers for AC/DC, film posters for Francis Ford Coppola and covers for *Time* magazine. But, like Warhol before him, he progressively detached himself from commercial work to pursue a career as a fine artist.

Schorr characterizes his paintings as an American variety of surrealism. His heroes are Salvador Dalí, that most technically polished of surrealists, and the Renaissance maestro of

bizarre organic inventions, Jerome Bosch. The competition is therefore tough, both technically and inventively.

How far does Schorr's loving transposition of lowbrow modes into art for galleries transcend his 'vulgar' sources? The professional illustrators of comics, pulp fiction and animated cartoons did their work with precise skills and superb draftsmanship. Looking at Schorr's retrospective exhibition in the San Jose Museum of Art in California, there are times when it seems

A laughing Mohican is dressed in the uniform of an Atlanta Braves baseball player. The palaeolithic Willendorf Venus dances with a stubbly and lecherous Mickey Mouse on a disco platform cut from the bloody leg of an ungulate. Another mouse in the ape's grip, presumably Minnie, plays the part performed so tellingly by Fay Wray in the 1933 film *King Kong*. Fronds of primitive equisetia spiral upwards into a vortex that is reminiscent of Paul Signac's pointillist portrait of

Félix Fénéon from 1890 (at the Museum of Modern Art in New York).

At one level, Schorr is revelling in the look of lowbrow culture. Yet once this is portrayed on the heroic scale of a historical painting on the walls of an art gallery, we are invited to adopt an ironic manner of viewing. Like all pop artists, he is having his cake and eating it too, only to spit it out at the end. We may intuit that the hunter-gatherer in Schorr's painting suggests that the apeman's primitive instinct for acquiring material trivia still drives our consumer culture today.

However, to formulate too fixed an interpretation is wrong. Unlike a

frame from a comic-book cartoon, the narrative is not spelled out by successive images and strings of words. The painting acts as an inviting field for interpretation.

Will Schorr be regarded as a major visual allegorist of our age, as was Botticelli in the Renaissance? For my part, I find it difficult to get past the unpleasantly garish rendering and unsubtle storytelling. But that is the point. The lowbrow artists are attuned to a major facet of our visual culture. And they are very Californian.

Martin Kemp is emeritus professor in history of art at the University of Oxford, Oxford, UK.



Todd Schorr's *Hunter Gatherer* depicts cartoon icons that are rich in surrealist symbolism.

that he is achieving little more than a collage of pop motifs in a stock surrealistic mode.

On other occasions, however, the narrative and symbolism move to a different level. This is particularly true of his recent series of ape paintings, which range across chronological territories from the remote past to the sci-fi future. *Hunter Gatherer* (owned by Leonardo di Caprio) garishly exemplifies how Schorr tells his stories and enriches them with intricate symbolism. We need to decode the whole and the parts in much the same way that a historical iconographer would tackle a Botticelli painting on a classical theme.

A humanoid ape emerges from a bubbling swamp, apparently amazed by the sight of a toy space robot. On his back, like Santa Claus, he carries a roughly stitched sack of cartoon character toys, including Batman.

Todd Schorr: American Surreal
San Jose Museum of Art, California
Until 16 September

NEWS & VIEWS



The Andromeda galaxy.

ASTROPHYSICS

Hidden chaos in cosmic order

Nickolay Y. Gnedin

“Galaxies, like elephants, have long memories,” says an influential article from the 1980s. Tapping into these memories has revealed some surprising facts about the history of our neighbouring Andromeda galaxy.

In popular literature and undergraduate textbooks, galaxies are often called the building blocks of the Universe. That implies that galaxies are complete, indivisible, monolithic blocks that make up all other cosmic structures. Indeed, for a long time that view was shared by many astronomers. After all, most of the galaxies that we see in books and on the Internet — and even in Chicago’s O’Hare International Airport, which celebrated the 2009 International Year of Astronomy by decorating its passageways with astronomical images — are perfect in the symmetries of their shapes. This respect for the strict beauty of galactic shapes is, to a large degree, due to Edwin Hubble. His classification of galaxies¹ on the basis of their shapes and structural properties is still in wide use today. But the perfection is never complete, and Hubble had to dump into a class of ‘irregular’ galaxies all the exceptions that did not fit into his two main classes of perfect ellipsoids or spiral disks. These irregular galaxies are often in the process of collision, or ‘merger’, with other galaxies. Galaxies are large and complex systems, and when they collide, everything becomes a mess.

Just like architects who enjoy the strict beauty of new buildings and disdain the ugliness of

ruins, many twentieth-century astronomers expressed their contempt for that (actually not-so-small) class of irregular galaxies by calling them train wrecks. But as telescopes became bigger and astronomical cameras more sensitive, several astronomers found themselves in the role of archaeologists rather than architects — to the point of even adopting the word ‘archaeology’ in their studies. On page 66 of this issue, McConnachie *et al.*² present the Pan-Andromeda Archaeological Survey (PAndAS), an excellent example of galactic archaeology in our own cosmic backyard. This large group of astronomers undertook a careful and painstaking study using the 3.6-metre Canada–France–Hawaii Telescope to map with unprecedented sensitivity (or depth, as astronomers like to call it) a large region around the Andromeda galaxy, a sister and neighbour galaxy to our own Milky Way.

The Milky Way and Andromeda coexist in an otherwise rather undistinguished corner of the cosmos. The only two ‘first ladies’ of their neighbourhood, with about 100 billion stars each, they are surrounded by around 50 satellite galaxies that range from small (with a mere 2 billion stars) to truly tiny (with just 1,000 stars). Like true ladies, they need elbow

room, and are far enough apart that their gravitational forces do not disturb each other very much.

But let’s focus on Andromeda. Every morning, as I come to my office, I walk past a beautiful wall-sized image of Andromeda from the Sloan Digital Sky Survey (ref. 3). The grandeur and symmetry of Andromeda often reminds me of a Gothic cathedral. What is there for an archaeologist to look for? Everything, claims the PAndAS team². Just as Gothic cathedrals usually stand on layers of archaeological treasures, so do the cathedrals of the cosmos. In their ultra-deep images of Andromeda and its neighbourhood, the PAndAS team discovered² a wealth of information about Andromeda’s violent past and its relations with its most famous satellite galaxy, the Triangulum galaxy.

How did they do that? Here is your chance to see galactic archaeology in action. The ultra-deep images of the Triangulum galaxy uncovered a huge — but very disperse, and hence faint — cloud of stars surrounding the small and dense Triangulum. The stars are clearly associated with Triangulum itself, but what force could have hurled millions of stars into its outer reaches? The answer is

obvious: the only 800-pound gorilla (in fact, a 7×10^{12} -pound one) in the room is Andromeda; its gravitational force is so strong that even a 2-billion-star galaxy such as Triangulum cannot withstand it.

The existence of an extended stellar halo is an important clue that in the past Triangulum came too close to its massive neighbour. By matching supercomputer simulations with the observed deep images of Triangulum and Andromeda, the PAndAS team² was able to reconstruct the most likely past trajectory of Triangulum. According to these calculations, the momentous encounter took place about 2.5 billion years ago, when Triangulum came to within a 'mere' 40,000 parsecs (130,000 light years) of Andromeda. The massive gravity of Andromeda caused tides in Triangulum so enormous that millions of stars were hurled beyond Triangulum's edge.

But surely such a close passage of Triangulum with its 2 billion stars must have left traces in Andromeda. And, indeed, the ultra-deep observations of the PAndAS team² reveal a mild

warp and small disturbances in Andromeda's stellar disk exactly at the spot where the passage of Triangulum would have caused them. These clues are strong enough to hold in any court of law.

Alas, the beauty of cosmic cathedrals is only perfect to the near-sighted; look deeper, and under the veil of perfection you'll find the messy traces of their violent past. Every galaxy is an irregular one, if one only looks deep enough. As astronomers are learning how to read these traces, they appreciate more and more that, as Binney and May exclaimed in their article⁴, "galaxies, like elephants, have long memories".

Nickolay Y. Gnedin is in the Theoretical Astrophysics Group at Fermilab, Batavia, Illinois 60510, USA.

e-mail: gnedin@fnal.gov

1. Hubble, E. P. *Astrophys. J.* **64**, 321–369 (1926).
2. McConnachie, A. W. *et al. Nature* **461**, 66–69 (2009).
3. www.sdss.org
4. Binney, J. & May, A. *Mon. Not. R. Astron. Soc.* **218**, 743–760 (1986).

CANCER

The fat and the furious

Eyal Gottlieb

Evidence linking metabolic alterations to cancer progression is accumulating. It seems that cancer cells must sustain their energy production and remain well fed to survive detachment from their normal habitat.

Good cells perform their physiological activities only at the right time and in the right place. When removed from their natural surroundings they self-destruct — an event that is termed anoikis (Greek for homelessness). Not so cancer cells: they survive, invade other tissues and continue to grow in unfamiliar territories. On page 109 of this issue, Schafer *et al.*¹ show that cancer-inducing oncogenes may protect cells from anoikis by maintaining the cells' glucose consumption. Their work demonstrates that the death of normal 'away-from-home' cells is caused by starvation, and strongly connects cellular metabolism to cancer.

Anoikis occurs when cells detach from the basement membrane or the extracellular matrix, both of which provide them with survival signals². Originally, anoikis was thought to be executed by apoptosis — a programmed cell death set off by several different cues, which results in cell fragmentation and elimination. Yet it turns out that blocking apoptosis does not prevent anoikis because detached cells die anyway³. Moreover, another process, autophagy — in which the cell digests some of its own components — is observed in anoikis cells⁴. If autophagy runs its full course, cells kill themselves by self-consumption but, at first, autophagy may help cells to survive

starvation. This prompted Schafer *et al.*¹ to investigate whether starvation is a feature of anoikis and, indeed, they showed that, when detached from the extracellular matrix, breast epithelial cells reduce their glucose uptake and energy production.

Cancer cells have an unbridled capacity for proliferation and invasion. Thus, cancer cells escape anoikis or, viewed another way, anoikis prevents cancer. Several oncogenes are known to hinder anoikis² but the mechanisms by which they do this remain obscure. One oncogene, *ERBB2*, encodes the epidermal growth factor receptor³, a cell-surface protein that is activated in approximately 25% of breast cancers. On the basis of their data linking glucose uptake with cellular detachment, Schafer and colleagues¹ asked whether *ERBB2* expression might prevent anoikis in detached breast epithelial cells by affecting energy metabolism. They grew normal and *ERBB2*-overexpressing breast epithelial cells in a three-dimensional culture system that mimics normal mammary gland structure. The breast cells spontaneously form globules that are made up of a single layer of cells surrounded by a basement membrane. These globules reliably emulate mammary gland sacs (or acini), which usually have a hollow centre

because the inner cells that have detached from the external basement membrane have been eliminated by anoikis — an essential process for the formation of mammary gland acini⁵. By contrast, filling the centre of the acini with cells is a feature of breast cancer (Fig. 1).

Using this culture system, Schafer *et al.* show that *ERBB2* expression rescues detached cells from energy depletion by maintaining their glucose uptake, specifically by activating the cancer-inducing PI3K/AKT pathway¹. The authors' findings are in line with results from previous studies showing that the PI3K/AKT pathway activates glucose uptake⁶ and protects cancer cells from starvation⁷.

Once in the cell, glucose may be metabolized through several pathways, including glycolysis, in which it is broken down to pyruvate to generate ATP, the cell's energy currency, and NADH, a mediator of ATP production. Pyruvate is routed to the mitochondria where, in the presence of oxygen, it is metabolized to produce large amounts of ATP (Fig. 2). Although oxygen is essential for generating the maximum yield of energy from glucose breakdown, it can also fatally damage the cell by contributing to various forms of oxidative stress. Glucose helps to prevent this oxidative stress — for a small price in bioenergy, glucose bypasses the initial steps of glycolysis and enters the pentose phosphate pathway. This produces less ATP but generates NADPH, a powerful mediator of antioxidative reactions that protect cells from oxidative damage^{8,9} (Fig. 2).

In an interesting twist, Schafer *et al.* note that anoikis can be prevented in normal, detached, glucose-starved cells if they are given antioxidants, showing that it is increased oxidative stress rather than decreased glycolysis that induces rapid anoikis. Moreover, the authors show that oxidative stress caused by a lack of

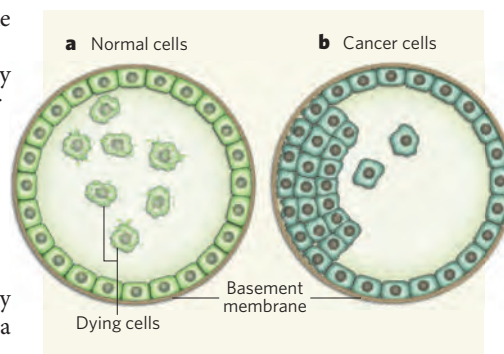


Figure 1 | Breast cancer cells survive detachment. Breast epithelial cells form hollow mammary gland sacs, termed acini. **a**, Normal breast epithelial cells are attached to a basement membrane, which provides them with survival signals. Cell death caused by detachment is called anoikis. **b**, *ERBB2*-overexpressing cancer cells survive detachment from the basement membrane and fill the lumen of the acinus. Schafer and colleagues¹ investigate the mechanisms of survival in breast cancer cells in a three-dimensional culture system that mimics normal mammary gland structure.

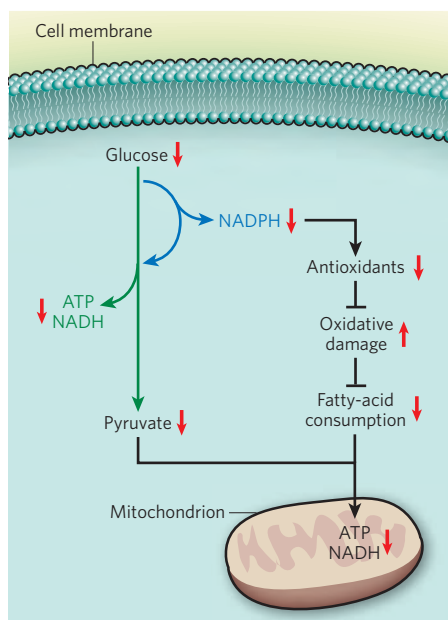


Figure 2 | Metabolic changes in detached normal breast epithelial cells. The red arrows represent the changes in the amounts of metabolites when normal breast epithelial cells detach from the extracellular matrix. Glucose uptake is decreased, which hinders both glycolysis (green) and the pentose phosphate pathway (blue). Consequently, an increase in oxidative damage blocks the consumption of fatty acids. The main energy-producing pathways are shut down and cells experience severe starvation and death. Schafer *et al.*¹ find that in detached ERBB2-overexpressing cancer cells, glucose uptake continues together with glycolysis and the pentose phosphate pathway. Sufficient energy is generated by glycolysis and by pyruvate breakdown in the mitochondria, and there is no need for compensatory fat consumption.

to take place? Perhaps anoikis is triggered by downregulation of the AKT protein. How does oxidative stress block the consumption of fats in normal detached cells? Does cellular starvation lead to finely tuned, regulated cell death? Or does it trigger an all-out offensive whereby all cell-death mechanisms are used if required? Apoptosis induction may be the pathway of choice for executing anoikis, but given that blocking apoptosis does not protect cells from starvation, there may well be alternative anoikis-inducing pathways that await discovery.

Eyal Gottlieb is at Cancer Research UK, Beatson Institute for Cancer Research, Switchback Road, Glasgow G61 1BD, UK.

e-mail: e.gottlieb@beatson.gla.ac.uk

1. Schafer, Z. T. *et al.* *Nature* **461**, 109–113 (2009).
2. Frisch, S. M. & Screaton, R. A. *Curr. Opin. Cell Biol.* **13**, 555–562 (2001).
3. Debnath, J. *et al.* *Cell* **111**, 29–40 (2002).
4. Lock, R. & Debnath, J. *Curr. Opin. Cell Biol.* **20**, 583–588 (2008).
5. Nelson, C. M. & Bissell, M. J. *Annu. Rev. Cell Dev. Biol.* **22**, 287–309 (2006).
6. Plas, D. R. & Thompson, C. B. *Oncogene* **24**, 7435–7442 (2005).
7. Kalaany, N. Y. & Sabatini, D. M. *Nature* **458**, 725–731 (2009).
8. Bensaad, K. *et al.* *Cell* **126**, 107–120 (2006).
9. Herrero-Mendez, A. *et al.* *Nature Cell Biol.* **11**, 747–752 (2009).
10. Tennant, D. A., Durán, R. V., Boulahbel, H. & Gottlieb, E. *Carcinogenesis* **30**, 1269–1280 (2009).

glucose uptake prevents the metabolism of fatty molecules (which are a compensating energy source) and so exacerbates starvation (Fig. 2).

Schafer and colleagues' observations¹ that ERBB2 overexpression sustains matrix-detached cancer cells by maintaining glucose uptake adds to the evidence connecting

metabolic alterations to cancer progression¹⁰. And their finding¹ that oxidative stress is a central mechanism that induces anoikis in matrix-detached cells calls into question the rationale of treating cancer with antioxidants. The authors' work raises many other intriguing questions. What shuts down glucose uptake and metabolism, allowing anoikis

parts from the aqueous environment.

But the polymer used by Eisenberg and colleagues consists of three parts: a hydrophilic poly(ethylene glycol) (PEG) block at one end; a hydrophobic polystyrene block in the middle; and poly(2-diethylaminoethyl methacrylate) (PDEA) at the other end. PDEA contains organic bases in its side chains, which harbour no electrical charges under basic conditions. As a result, PDEA is hydrophobic in basic solution, and occupies the inner part of the vesicle membrane (Fig. 1a, overleaf). However, in acidic conditions, the bases are protonated and PDEA becomes positively charged, making it hydrophilic. The PDEA block therefore starts to attract water, because it would actually prefer to be dissolved in water.

The polarization of the PDEA block in response to a change from basic to acidic conditions would normally cause the vesicle to break up. However, the simple but smart design of Eisenberg and colleagues' copolymer prevents the vesicle from disintegrating. At high pH, the non-polar polystyrene is partly mixed with the similarly non-polar PDEA (Fig. 1a). But at low pH, the polystyrene is forced out of the now positively charged PDEA layer because of the difference in polarity between the two polymers. The polystyrene then forms a specific region of its own between the PDEA and the PEG layers of the vesicle membrane (Fig. 1b). Because polystyrene is a glassy polymer, this region is rigid, and so prevents

MATERIALS SCIENCE

Pulsating vesicles

Jan C. M. van Hest

During her travels through Wonderland, Alice finds several ways of growing and shrinking in size. A polymeric vesicle plays the same trick in response to pH, in a process that might one day be useful for drug delivery.

Molecules and molecular assemblies that undergo controllable shape changes in response to an external trigger will be crucial to manufacturing nanometre-scale devices. But although stimulus-responsive systems are common in nature, it is not always easy to create synthetic versions, even when using nature for inspiration. This is certainly true of vesicles, a class of cell-mimicking structures that have been investigated intensively for decades. In the *Journal of the American Chemical Society*, Eisenberg and colleagues¹ report a polymeric vesicle for which the structure and properties can be controlled — it 'breathes' in response to pH, becoming twice as big in acid solution as it is in basic solution. This reversible size change is accompanied by a permeability change, so that water molecules diffuse more easily through the membrane wall of the vesicle when the structure is highly swollen.

Vesicles can be regarded as nanometre-

scale containers, in which a molecular shell or membrane encloses an aqueous interior. Their potential applications are mostly linked to biomedicine, and range from drug-delivery devices to nanovessels for chemical reactions. For all of these applications, the ability to control the transport of molecules in and out of the container is paramount; Eisenberg and colleagues' work reveals one way in which this can be done.

Their molecular assembly¹ belongs to a special class of vesicle known as polymerosomes². The membranes of most polymerosomes are made up of polymeric molecules that contain a water-soluble (hydrophilic) block and a water-insoluble (hydrophobic) block. These surfactant characteristics result in the spontaneous aggregation of the molecules in water into bilayers that form spherical shells; the hydrophilic parts of the molecules line the outer faces of the bilayer, fully shielding the hydrophobic

movement and diffusion of polymer blocks through the membrane. It therefore functions as a kind of molecular restraint that keeps all the loose polymeric components together and in order, thereby preventing PDEA from reorienting itself towards the water layer and so stopping the vesicle from irreversibly losing its shape. The only thing that the PDEA layer can do when it attracts water is swell, which causes the entire vesicle to swell in turn. Because the water-swollen layer is less densely packed than before, water molecules pass more easily through the polymer membrane.

Polymeric vesicles that change their size in response to pH have been described before. In one case³, a conformational change of the water-soluble block resulted in an increase in the vesicle's size. In another example⁴, the structure swelled because of protonation of a PDEA block (as in Eisenberg and colleagues' polymersome¹), but the vesicle structure was

maintained by cross-linking of the polymer molecules. The unique feature of the capsule reported by Eisenberg and colleagues is that it can 'breathe' repeatedly — swell and contract in response to pH changes — without the need for chemical cross-linking of the polymer molecules to stabilize its structure.

The authors' study¹ also demonstrates that there is still much to learn about the aggregation of macromolecules that contain hydrophilic and hydrophobic regions. Because such aggregations are often non-equilibrium processes, this makes them especially difficult to predict and control. Much more work is therefore required before these polymeric self-assembly processes can be used to make nanometre-scale devices. Moreover, polymeric aggregates are often stable structures, even when not in equilibrium, which means that many different aggregation morphologies might be obtainable from the same polymer, if the correct formation process is applied. This could lead to many exciting possibilities for objects such as nanowires.

Of course, if Eisenberg and colleagues' polymersome is to be useful in a biological setting, the vesicle's size must be controllable under physiological conditions, and its membrane must allow through not only water molecules, but also biologically active compounds. The second issue is already being tackled by an

increasing number of scientists, who have succeeded in making polymersomes that irreversibly trap enzymes, but which allow the enzyme's substrates and products to pass through the vesicle membrane. The selective permeability originates from the clever design of the polymer molecules⁵, or from trapping protein channels in the vesicle's membrane⁶.

Eisenberg and colleagues' work¹ breathes new life into the branch of polymer science that studies the aggregation phenomena of apparently simple block copolymers. By combining lessons from polymer science with those from nature, nanometre-scale device manufacturing will be elevated to a higher level. ■

Jan C. M. van Hest is in the Department of Organic Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, Heyendaalseweg 135, Nijmegen 6525 AJ, the Netherlands.
e-mail: j.vanhest@science.ru.nl

1. Yu, S., Azzam, T., Rouiller, I. & Eisenberg, A. *J. Am. Chem. Soc.* **131**, 10557–10566 (2009).
2. Discher, D. E. & Eisenberg, A. *Science* **297**, 967–973 (2002).
3. Chécot, F., Lecommandoux, S., Gnanou, Y. & Klok, H.-A. *Angew. Chem. Int. Edn* **41**, 1339–1343 (2002).
4. Du, J. & Armes, S. P. *J. Am. Chem. Soc.* **127**, 12800–12801 (2005).
5. Kim, K. T., Cornelissen, J. J. L. M., Nolte, R. J. M. & van Hest, J. C. M. *Adv. Mater.* **21**, 2787–2791 (2009).
6. Nardin, C., Thoeni, S., Widmer, J., Winterhalter, M. & Meier, W. *Chem. Commun.* 1433–1434 (2000).

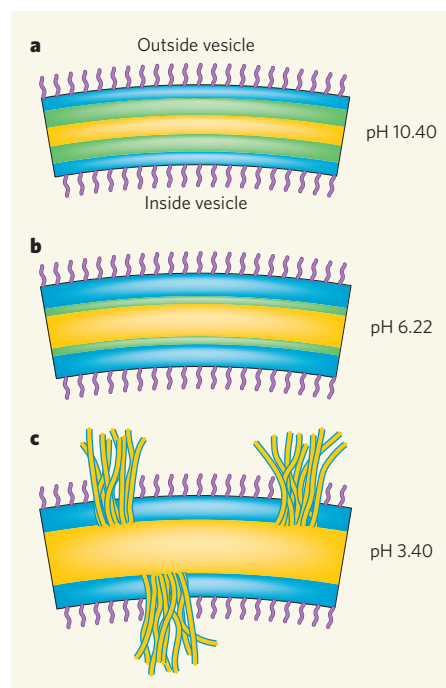


Figure 1 | A pH-responsive vesicle membrane. Eisenberg and colleagues¹ report polymeric vesicles that swell and shrink in response to changes in pH. The molecules that make up the membrane bilayer of the vesicles consist of one end block of poly(ethylene glycol) (PEG); a middle block of polystyrene; and a block of poly(2-diethylaminoethyl methacrylate) (PDEA) at the other end. **a**, At high pH, the pattern of polymer layers in the membrane is as follows: an outer layer of PEG (purple), a thin polystyrene layer (blue), a thicker polystyrene-PDEA layer (green), then a PDEA layer (yellow) at the centre of the membrane; the reverse pattern occurs moving from the PDEA layer to the inside of the vesicle. **b**, As the pH decreases, the PDEA and polystyrene layers separate, each becoming thicker, whereas the polystyrene-PDEA layers become thinner. **c**, At low pH, the polystyrene layer might rupture in places, allowing some PDEA to break through.

STRUCTURAL BIOLOGY

A channel with a twist

Valeria Vásquez and Eduardo Perozo

Mechanosensitive channels release tension in cell membranes by opening 'pressure relief' pores. The structure of a partially open channel suggests a gating mechanism and delivers an unexpected architectural twist.

In the constant struggle with their environment, free-living cells have evolved a variety of mechanisms to deal with sudden variations in the physicochemical properties of their surroundings¹. But few environmental challenges require more assertive responses than a hypo-osmotic insult. Faced with a sudden decrease in osmotic pressure, as might be caused by a spring downpour, free-living cells are subjected to a rapid influx of water. This sharply increases the pressure of the cell contents against the membrane, potentially compromising the integrity of the cell.

Most prokaryotes (bacteria and archaea) have therefore evolved a 'pressure-release valve' mechanism in which changes in membrane tension open up channels to form large, aqueous pores in the membrane. Once formed, the short-lived pores allow the passage of both solute and solvent at very high rates, quickly equilibrating hypo-osmotic imbalances

across the cell wall. Different variants of these mechanosensitive (MS) channels exist, each of which forms pores that have distinct conductance and tension thresholds^{2–5}. On page 120 of this issue, Rees and colleagues⁶ provide evidence for the gating mechanism in the mechanosensitive channel of large conductance of *Staphylococcus aureus* (SaMscL) by determining the crystal structure of a truncated form of the channel.

To solve the molecular mechanism of gating in prokaryotic MS channels, two crucial questions must be addressed: how are membrane deformations sensed by the channel, and what is the mechanism by which the channel physically opens its permeation pathway? Although we have some idea of the types of force at the membrane-channel interface that might trigger MS channel opening⁷, the principles underlying tension sensing through the membrane remain largely unknown. We do, however,



50 YEARS AGO

The Human Response to an Expanding Universe. By Harlow Shapley — [The author of this book] is a world-renowned figure in the fields of astronomy and cosmography... Dr. Shapley begins by attempting an obituary of the anthropocentric view that man is the centre of the cosmos, and continues by interpreting the consequences to man (or rather to certain aspects of rational thinking) of the latest scientific discoveries in the cosmos...

Dr. Shapley's displacement of human life from its once supreme position does not make him a pessimist, for he argues cogently that there must be at least a hundred million planets capable of supporting some form of life. Dr. Shapley concludes his book with what he calls "a Martian look" into the future. He dismisses the prospect of the Earth's collision with a star, or of wandering from its orbit and getting too near or too far away from the Sun. Nor does he envisage a biological calamity wiping out the whole human race. The real danger is man himself, who is busy perfecting the tools for performing an operation which is unlikely to be performed by natural forces.

From *Nature* 5 September 1959

100 YEARS AGO

In February last Dr. N. Annandale obtained on the Orissa coast of India a number of small more or less nearly globular organisms in the tide-wash. When placed in water their shape changed from globular to conical, and indicated that they were evidently pelagic sea-anemones, although devoid of tentacles. The mouth is conspicuous, forming a relatively long, narrow slit expanded at one end, and the whole organism presents a milky appearance... As these actinians, which are apparently adult, although no gonads are visible, evidently indicate a new generic and specific type, Dr. Annandale has described them under the name *Anactinia pelagica*.

From *Nature* 2 September 1909

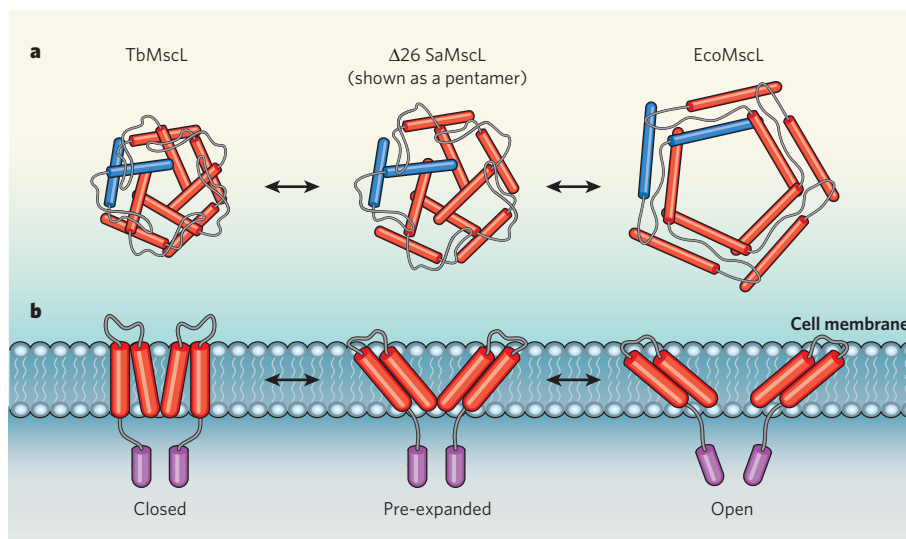


Figure 1 | Opening of the MscL channel. The MscL channel opens in cell membranes to release osmotic pressure in the cell. The crystal structure of the MscL from *Mycobacterium tuberculosis*⁸ (TbMscL) revealed the closed state of the channel, whereas the open state was seen in the spectroscopically derived structure¹⁰ of EcoMscL (the MscL from *Escherichia coli*). Rees and colleagues⁶ now report the crystal structure of an intermediate, 'pre-expanded' state of a truncated MscL from *Staphylococcus aureus* (Δ26 SaMscL). **a**, In these extracellular views looking down at the MscL structure, the movements of the helices that line the pore as the channel opens (or closes) are visible. The cross-section of the channel expands from left to right, but the pore remains closed in the pre-expanded state. The Δ26 SaMscL channel is actually a tetramer, but is shown here as a pentamer for comparison with the open and closed states. An individual monomer is highlighted in blue. **b**, Side views reveal that the helices lining the inside of the pore tilt towards the plane of the membrane as the channel moves from the closed to the pre-expanded state, and that the whole channel structure becomes flatter. The helices maintain this orientation as the cross-section of the channel expands to reach the open state.

have a better handle on the conformational rearrangements in the channels that underlie their opening.

MscL is nonselective for the ions and small molecules it transports and is activated at membrane tensions close to the breaking point of the membrane bilayer. It is thus usually thought of as a bacterium's last line of defence against hypo-osmotic shock. Over the past decade, genetic, structural and biophysical data have provided a fairly detailed picture of the channel's basic architecture⁸ and functional behaviour⁹, and have defined the types of structural rearrangement that could support the formation of very large pores characteristic of MscL^{10,11}. Rees's group previously reported the MscL crystal structure from *Mycobacterium tuberculosis*⁸ (TbMscL), which is generally agreed to represent the closed conformation of the channel. The new structure⁶ looks like an intermediate conformation, somewhere between the open and closed states.

It has been proposed^{12,13} that a domain of MscL known as the cytoplasmic bundle acts as a sieve that limits the passage of large molecules through the pore. By removing the last 26 amino-acid residues of this cytoplasmic domain from SaMscL, Rees and colleagues⁶ were able to drive the (typically stable) closed conformation of the channel to an expanded, partially open state. This state is characterized by a significant tilt of the TM1 helices —

transmembrane helices that line the pore of the channel — towards the plane of the membrane. The conformational change shortens the length of the water-permeation pathway and makes the whole channel flatter than it is in the closed state (Fig. 1). The authors also found that the activity of, and the current through, their truncated channels in functional measurements is greater than that of wild-type channels, supporting the idea that the structural changes observed in the crystal structure also occur in functional channels.

Rees and colleagues' crystal structure⁶ provides an explicit conformational pathway from the closed to the open state of SaMscL: the first physical transition generates a 'pre-expanded' state in which the cross-sectional area (*A*) of the channel is slightly larger than in the closed state (Fig. 1). This is likely to be the most tension-dependent state of the mechanism, as the probability that the channel will open should be proportional to $-\gamma\Delta A$, where γ is the lateral tension in the membrane⁹. The pre-expanded state maintains a narrow pore, still flanked by the inner TM1 helices, and is therefore predicted to be non-conductive. The present structure⁶ is thus in excellent agreement with earlier models of the pre-expanded state^{9,11,14}.

Although Rees and colleagues' structure does not have a pore wide enough to conduct, the fact that the transmembrane helices are tilted away from the normal of the membrane, together with the expansion of the

cross-sectional area of the pore, clearly suggests that the observed conformation is an intermediate stage on its way to the open state. Indeed, the observed tilts of the helices match remarkably well with those predicted by spectroscopic studies of channels trapped in the open state¹⁰.

Perhaps the most surprising result of this work⁶ is that, defying all expectations, the channel is arranged as a tetramer, not a pentamer as observed for the crystal structure of TbmScL in a dosed state⁸. This is in contrast to the vast majority of oligomeric membrane proteins, each of which maintains the same oligomeric architecture across species and physiological states, at least for those studied so far. Still, controversy regarding the oligomeric state of MscL is not new — it was originally thought to be a hexamer, on the basis of biochemical experiments¹⁵ and electron-crystallography data¹⁶. It is, however, unlikely that the present result⁶ will lend further support to the idea that MscL gating is the consequence of monomer aggregation in the plane of the bilayer (which could support variable oligomeric states).

The observed tetrameric structure of truncated SaMscL raises several questions. Does the oligomeric state of MscL vary in different species? Does the channel reside in the membrane as different multimers? If so, does each multimer have different functional properties? And are particular multimers favoured by particular physiological conditions? Regardless of the answers, the plasticity of this family of ion channels seems to be exceptional. More

time and experimental insight will surely be needed to realize the full implications of this structure⁶.

Valeria Vásquez is in the Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. Eduardo Perozo is in the Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA.

e-mail: eperozo@uchicago.edu

1. Broder, S. & Venter, J. C. *Annu. Rev. Pharmacol. Toxicol.* **40**, 97–132 (2000).
2. Anishkin, A. & Kung, C. *Curr. Opin. Neurobiol.* **15**, 397–405 (2005).
3. Booth, I. R., Edwards, M. D., Black, S., Schumann, U. & Miller, S. *Nature Rev. Microbiol.* **5**, 431–440 (2007).
4. Kung, C. *Nature* **436**, 647–654 (2005).
5. Martinac, B., Saimi, Y. & Kung, C. *Physiol. Rev.* **88**, 1449–1490 (2008).
6. Liu, Z., Gandhi, C. S. & Rees, D. C. *Nature* **461**, 120–124 (2009).
7. Wiggins, P. & Phillips, R. *Biophys. J.* **88**, 880–902 (2005).
8. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. *Science* **282**, 2220–2226 (1998).
9. Sukharev, S. I., Sigurdson, W. J., Kung, C. & Sachs, F. *J. Gen. Physiol.* **113**, 525–540 (1999).
10. Perozo, E., Cortes, D. M., Sompornpisut, P., Kloda, A. & Martinac, B. *Nature* **418**, 942–948 (2002).
11. Sukharev, S., Betanzos, M., Chiang, C. S. & Guy, H. R. *Nature* **409**, 720–724 (2001).
12. Cruickshank, C. C., Minchin, R. F., Le Dain, A. C. & Martinac, B. *Biophys. J.* **73**, 1925–1931 (1997).
13. van den Bogaart, G., Krasnikov, V. & Poolman, B. *Biophys. J.* **92**, 1233–1240 (2007).
14. Betanzos, M., Chiang, C. S., Guy, H. R. & Sukharev, S. *Nature Struct. Biol.* **9**, 704–710 (2002).
15. Sukharev, S. I., Schroeder, M. J. & McCaslin, D. R. *J. Membr. Biol.* **171**, 183–193 (1999).
16. Saint, N. et al. *J. Biol. Chem.* **273**, 14667–14670 (1998).

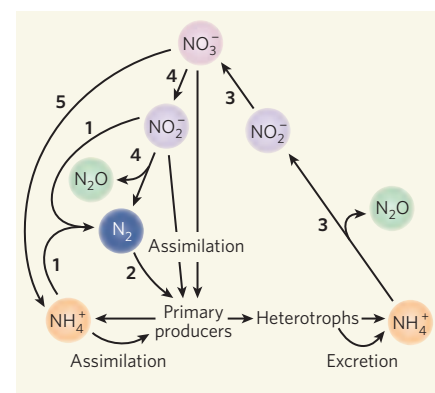


Figure 1 | Principal features of the marine nitrogen cycle. 1, Anammox, involving the generation of N_2 from inorganic constituents by autotrophic microbes. 2, Nitrogen fixation. 3, Nitrification. 4, Heterotrophic denitrification, in which N_2 is produced from NO_3^- used in microbial respiration of organic substrates; this process also results in the production of NH_4^+ and CO_2 (not shown). 5, Dissimilatory nitrate reduction to ammonium (DNRA). NO_3^- , nitrate; NO_2^- , nitrite; N_2O , nitrous oxide; NH_4^+ , ammonium. The N_2 -output pathways, 1 and 4, were the subject of Ward and colleagues' study³.

respiration; this was long viewed as the main metabolic pathway for N_2 production in midwater OMZs. But the picture was complicated by the discovery that another route — anammox, a pathway that produces N_2 by coupling the reduction of nitrite to the oxidation of ammonium — has an important role in the oceanic nitrogen cycle^{4,5}. The autotrophic microbes in this pathway produce biomass from inorganic molecules by harvesting energy from the anammox reaction. Both pathways are shown in Figure 1, which summarizes the overall marine nitrogen cycle.

Studies in the Black Sea⁵, the Benguela upwelling system⁶ and the ETSP⁷ suggested that anammox may in fact be the dominant pathway removing reactive nitrogen from the ocean. If so, however, what supplies the necessary substrates? Heterotrophic denitrification can potentially do so — it can supply both nitrite, an intermediate in the denitrification pathway, and ammonium, a product of heterotrophy, to support anammox⁴. Earlier this year, however, Lam et al.⁷ suggested a major revision of the nitrogen cycle in which a process called dissimilatory nitrate reduction to ammonium (DNRA) provides the necessary ammonium for anammox. In this revised cycle, heterotrophic denitrification has a minor role and is no longer a significant source of N_2 .

Part of the uncertainty surrounding heterotrophic denitrification is due to the difficulty of measuring the rates of the processes involved. Even low levels of oxygen contamination suppress the activity of denitrifying microbes, and the high concentration of N_2 in the water greatly reduces the sensitivity of experimental methods that are based on following the movement of the ^{15}N isotope tracer into the N_2 pool.

NITROGEN CYCLE

Oceans apart

Maren Voss and Joseph P. Montoya

Reactive nitrogen is lost from the oceans as dinitrogen — N_2 — produced by microbial metabolism. The latest twist in an ongoing story is that different pathways dominate in two of the oceanic regions concerned.

The availability of nitrogen limits biological production in much of the world ocean¹; this in turn affects the strength of the 'biological pump' that converts carbon dioxide into organic matter that can sink and be sequestered in the deep sea². The main input to the marine nitrogen cycle comes from the fixation of nitrogen gas (N_2) into biologically available forms. The main output is through biological processes that generate a return flux of N_2 . Both input and output remain poorly constrained³, and there is a pressing need to define them better.

On page 78 of this issue, Ward et al.³ describe how they have investigated the output side of the budget. They report on a comprehensive field study aimed at identifying the processes

and organisms responsible for N_2 production in two of the major sites of nitrogen loss in the world ocean. The sites concerned are the oxygen minimum zones (OMZs) of the Arabian Sea and the Eastern Tropical South Pacific (ETSP). At these sites, microbial degradation of sinking organic matter obtained by primary production — mainly photosynthesis — in the surface ocean completely removes oxygen from large parts of the water column. The resulting conditions favour metabolic pathways that convert nitrogen from its biologically reactive forms (for example, nitrate and ammonium) to N_2 .

In one of the N_2 -conversion pathways, termed heterotrophic denitrification, nitrate acts as a terminal electron acceptor in microbial

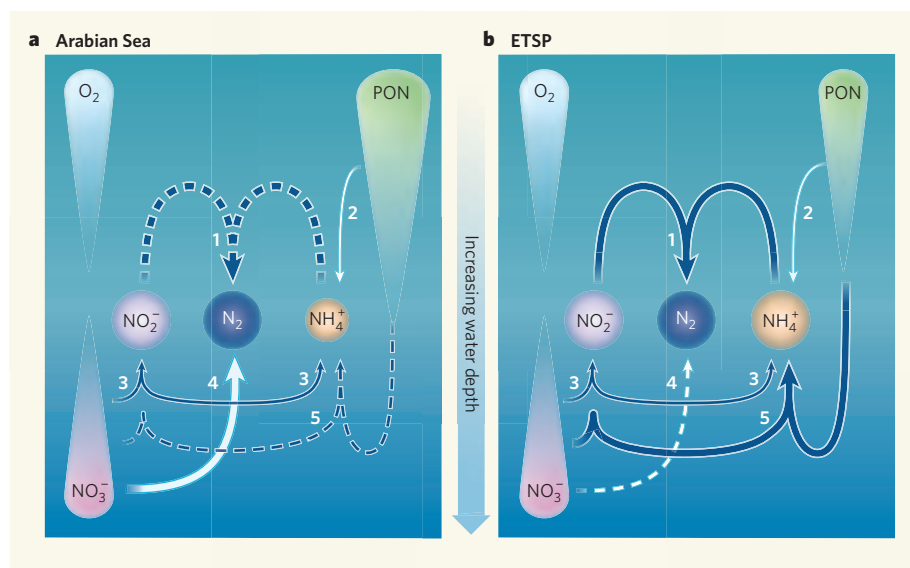


Figure 2 | Two oceans, two nitrogen cycles. Outlines of nitrogen cycling in the oxygen minimum zones of (a) the Arabian Sea and (b) the Eastern Tropical South Pacific (ETSP). Teardrops show typical profiles of concentration as a function of depth; PON, particulate organic nitrogen. Blue lines denote pathways related to the anammox–DNRA cycle; white lines are related to heterotrophic denitrification. 1, Anammox; 2, remineralization; 3, nitrate reduction; 4, heterotrophic denitrification; 5, DNRA. As Ward *et al.* show³, in the Arabian Sea denitrification is much more important than anammox in generating loss of nitrogen from the system as N_2 . In the ETSP, anammox is fuelled by ammonium from DNRA and is an important route to N_2 .

Ward *et al.*³ have overcome these difficulties by carrying out a set of experiments designed to evaluate the relative prevalence of heterotrophic denitrification and anammox.

Their main result is that the DNRA–anammox nitrogen cycle does not have a major role in the Arabian Sea (Fig. 2a), the largest of the world's three OMZs. Ward *et al.* show convincingly that it is indeed heterotrophic denitrification that accounts for most of the nitrogen losses in this system. By contrast, they measured much lower rates of denitrification in the ETSP (Fig. 2b), a finding consistent with the core argument of Lam and colleagues⁷. In both the Arabian Sea and the ETSP, however, the authors' molecular data³ showed the presence of heterotrophic denitrifiers in the water column at much higher abundances than the anammox organisms. This raises the obvious questions of why the strength of heterotrophic denitrification differs so much between the two OMZs, and which factors promote denitrifier activity in the Arabian Sea and not in the ETSP.

There is now strong evidence that heterotrophic denitrification does have a central role in removing reactive nitrogen from the ocean. But we still lack a clear understanding of why it is so much more important in the Arabian Sea. Ward *et al.*³ suggest that a difference in the magnitude and timing of the supply of organic matter to the OMZ is the critical factor. Indeed, the Arabian Sea seems to be more productive on average⁸, and it is known to support large blooms of nitrogen-fixing organisms⁹ that introduce extra biomass into the surface waters. By contrast, production in the ETSP may be more episodic. During periods of increased delivery

of organic matter to the OMZ, processes in the ETSP may more closely resemble those in the Arabian Sea, a possibility that can be tested only through additional studies at sea.

The workings of the oceanic nitrogen cycle may seem arcane, but understanding them is a necessary component in understanding both global marine productivity and climate change. The two major pathways to N_2 — heterotrophic denitrification and anammox — each reduce the ocean's stock of reactive nitrogen, but with

different collateral effects. Denitrification is a significant source of nitrous oxide (N_2O)¹⁰, a greenhouse gas, and its heterotrophic nature makes it a potential short-circuit in the biological pump that transfers carbon from the atmosphere into the deep sea. By contrast, anammox produces no greenhouse gases, and is an autotrophic process that can potentially increase the efficiency of the biological pump by reducing the net production of CO_2 within the water column. Shifts in the productivity of the oceans due to rising temperatures and levels of CO_2 are likely to affect the distribution and extent of OMZs¹¹. Knowing which nitrogen-cycle processes are at work, and where, will be an essential aspect of gauging the likely response of the oceans to global-scale changes in temperature and CO_2 .

Maren Voss is at the Leibniz Institute for Baltic Sea Research, Warnemünde, Seestrasse 15, D-18119 Rostock, Germany. Joseph P. Montoya is in the School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. e-mails: maren.voss@io-warnemuende.de; joseph.montoya@biology.gatech.edu

1. Codispoti, L. A. *Biogeochemistry* **4**, 233–253 (2007).
2. Sabine, C. L. *et al. Science* **305**, 367–371 (2004).
3. Ward, B. B. *et al. Nature* **461**, 78–81 (2009).
4. Dalsgaard, T. *et al. Nature* **422**, 606–608 (2003).
5. Kuypers, M. M. M. *et al. Nature* **422**, 608–611 (2003).
6. Kuypers, M. M. M. *et al. Proc. Natl Acad. Sci. USA* **102**, 6478–6483 (2005).
7. Lam, P. *et al. Proc. Natl Acad. Sci. USA* **106**, 4752–4757 (2009).
8. Longhurst, A. *et al. J. Plankton Res.* **17**, 1245–1271 (1995).
9. Capone, D. G. *et al. Mar. Ecol. Progr. Ser.* **172**, 281–292 (1998).
10. Bange, H. in *Nitrogen in the Marine Environment* (eds Capone, D. *et al.*) 51–94 (Elsevier, 2008).
11. Fabry, V. J. *et al. ICES J. Mar. Sci.* **65**, 414–432 (2008).

NEUROSCIENCE

Persistent feedback

Hyojung Seo and Daeyeol Lee

How does the brain remember the consequences of our actions? Persistent activity in the prefrontal cortex and basal ganglia may be crucial for learning correct actions through experience.

Do you jump out of bed when you hear the alarm clock ring in the morning? Or do you push the snoozer? Your choice will depend on the consequences of similar actions in the past. Typically, if an action triggered by a stimulus leads to a pleasant outcome, such as food or safety, we are more likely to perform the same action on re-encountering the same stimulus¹. Therefore, a fundamental building block in shaping behaviour is the relationship between a sensory event, a chosen action and its consequences, but how the brain stores this information is still a matter of speculation.

A recent paper in *Neuron* by Histed *et al.*² sheds some light on these mechanisms by showing that neurons in the primate prefrontal cortex and basal ganglia display persistent activity that is related to the outcomes of previous actions.

The task of forming an association between a particular sensory event, a behavioural response and the outcome of this response (Fig. 1) is not trivial, as a particular outcome might be preceded by multiple sensory stimuli and responses. To discover which of these events are related, it is often necessary to

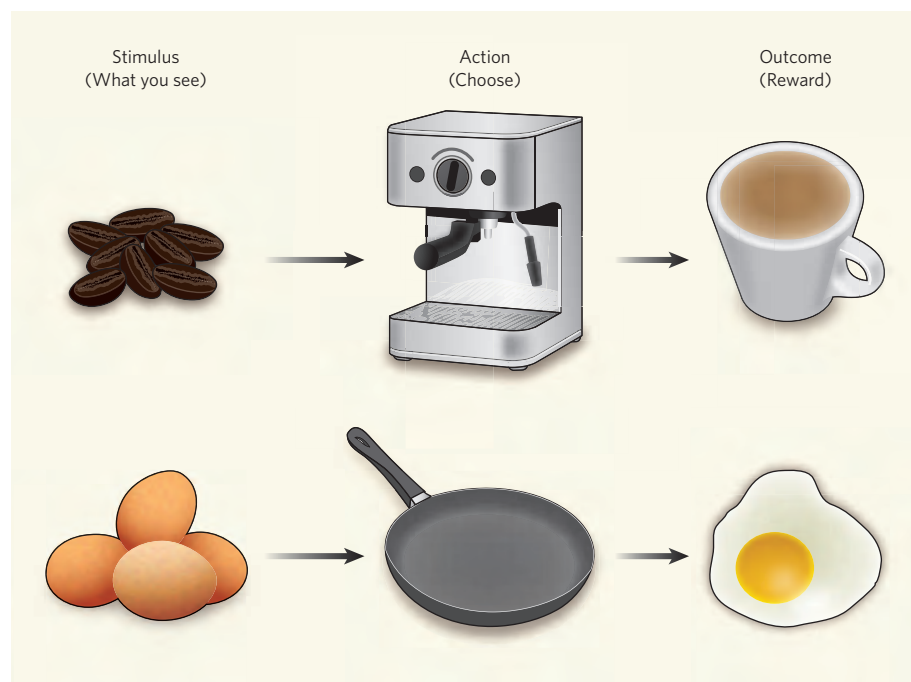


Figure 1 | Stimulus-action-outcome associations. Histed and colleagues² study how the brain stores information about the relationship between a stimulus, an action and its outcome. Their work shows that learning the outcomes of specific actions might be facilitated by persistent neural activity in the prefrontal cortex and basal ganglia.

experience the same sequence of sensory stimuli, actions and outcomes repeatedly. Moreover, the sensory and motor events that need to be remembered together are often transient, whereas the outcome of an action may be revealed only after a long delay. Therefore, the information about various sensory and motor events must be stored temporarily before the long-lasting memory about their relationship can be formed.

At least two mechanisms have been proposed to store such information in the brain. First, the relationship between two events, such as the tendency for separate sensory stimuli to occur together, might be reflected in the strength of synapses within a population of neurons^{3,4}. Although the properties of such synaptic plasticity are becoming better characterized, how they encode behaviourally relevant information during the learning of a specific task is poorly understood. In addition, this mechanism requires the activity of the pre-synaptic and post-synaptic neurons to overlap within a relatively narrow time window, and therefore might not be able to deal with a long delay between relevant events³. A second, alternative mechanism for storing information about the relationship between multiple events might be persistent neural activity, which is sustained beyond the duration of the initiating event^{5,6}.

The best-known example of persistent neural activity is the 'delay activity' observed during the delay period of a working-memory task^{5,6}. The activity of neurons in the primate prefrontal cortex often alters while the animal remembers the location or identity of a partic-

ular stimulus in order to use this information for guiding subsequent behaviour. The results from more recent studies⁷⁻¹¹, however, suggest that persistent activity might serve broader functions than working memory. For example, persistent activity is observed⁷ in the prefrontal cortex, even when the animal is not performing a working-memory task. Also, during a dynamic decision-making task in which the animal has to discover an optimal behavioural strategy by trial and error, the activity of neurons in the prefrontal cortex and parietal cortex often changes according to the outcomes of the animal's previous choices⁸⁻¹¹.

Histed and colleagues' work² shows that persistent neural activity might have an important role in learning the correct actions, even when the animal needs to repeatedly revise the associations between a sensory stimulus, an action and its outcome. In their study, monkeys learned which of two alternative actions (looking to the left or to the right) was required to gain a reward after viewing a particular visual stimulus. Once the animal learned the rewarded or 'correct' actions, the associations between stimuli and the correct actions were unpredictably reversed, requiring the animal to re-learn the relationship. The authors found that individual neurons in the prefrontal cortex and basal ganglia showed changes in activity depending on whether the action was correct or incorrect. And these neurons often displayed persistent activity during the inter-trial interval, signalling whether the animal had performed correctly or not in the preceding trial. Histed *et al.* also found that the activity in the prefrontal cortex

and basal ganglia that is related to the animal's upcoming choice was enhanced, and the animal was more likely to choose the correct action when its choice in the previous trial was correct.

Histed and colleagues' results indicate that persistent neural activity might have more diverse roles in relating multiple events separated in time than previously thought. Studies have shown that some neurons in the prefrontal cortex and basal ganglia carry information about previous actions^{8,11-13}. Combined with the authors' findings², these data suggest that the prefrontal cortex and basal ganglia might be essential brain areas for storing information about action-outcome associations¹²⁻¹⁴. However, how the signals related to an action and its outcome can be bound to a particular stimulus remains unknown. For example, in the task used by Histed *et al.*, the animal had to learn the relationship between a particular stimulus and an action. But neurons in the prefrontal cortex and basal ganglia predominantly encoded the outcome of the animal's choice between trials rather than the corresponding stimulus or action. To build on these findings, more studies are needed to further characterize the neural circuitry responsible for forming appropriate associations necessary for adaptive and flexible behaviours.

Hyojung Seo and Daeyeol Lee are in the Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.
e-mail: daeyeol.lee@yale.edu

- Thorndike, E. L. *Animal Intelligence* (MacMillan, 1911).
- Histed, M. H., Pasupathy, A. & Miller, E. K. *Neuron* **63**, 244-253 (2009).
- Wickens, J. R. *Behav. Brain Res.* **199**, 119-128 (2009).
- Fusi, S., Asaad, W. F., Miller, E. K. & Wang, X. J. *Neuron* **54**, 319-333 (2007).
- Fuster, J. M. & Alexander, G. E. *Science* **173**, 652-654 (1971).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. *J. Neurophysiol.* **61**, 331-349 (1989).
- Meyer, T., Qi, X.-L. & Constantinidis, C. *Cereb. Cortex* **17**, i70-i76 (2007).
- Barracough, D. J., Conroy, M. L. & Lee, D. *Nature Neurosci.* **7**, 404-410 (2004).
- Seo, H. & Lee, D. *J. Neurosci.* **27**, 8366-8377 (2007).
- Narayanan, N. & Laubach, M. *J. Neurophysiol.* **100**, 520-525 (2008).
- Seo, H., Barracough, D. J. & Lee, D. *J. Neurosci.* **29**, 7278-7289 (2009).
- Kim, Y. B. *et al.* *J. Neurophysiol.* **98**, 3548-3556 (2007).
- Ito, K. & Doya, K. *J. Neurosci.* **29**, 9861-9874 (2009).
- Rudebeck, P. H. *et al.* *J. Neurosci.* **28**, 13775-13785 (2008).

Correction

The News & Views article "Chemical physics: Electronic movies" by Marc Vrakking (*Nature* **460**, 960-961, 2009) stated at the end of the fifth paragraph that "A similar conclusion was recently reached in a study of harmonic generation from nitrogen molecules", and incorrectly cited reference 8 of the article. This should have cited B. K. McFarland *et al.* *Science* **322**, 1232-1235 (2008).

REVIEWS

Early-warning signals for critical transitions

Marten Scheffer¹, Jordi Bascompte², William A. Brock³, Victor Brovkin⁵, Stephen R. Carpenter⁴, Vasilis Dakos¹, Hermann Held⁶, Egbert H. van Nes¹, Max Rietkerk⁷ & George Sugihara⁸

Complex dynamical systems, ranging from ecosystems to financial markets and the climate, can have tipping points at which a sudden shift to a contrasting dynamical regime may occur. Although predicting such critical points before they are reached is extremely difficult, work in different scientific fields is now suggesting the existence of generic early-warning signals that may indicate for a wide class of systems if a critical threshold is approaching.

It is becoming increasingly clear that many complex systems have critical thresholds—so-called tipping points—at which the system shifts abruptly from one state to another. In medicine, we have spontaneous systemic failures such as asthma attacks¹ or epileptic seizures^{2,3}; in global finance, there is concern about systemic market crashes^{4,5}; in the Earth system, abrupt shifts in ocean circulation or climate may occur⁶; and catastrophic shifts in rangelands, fish populations or wildlife populations may threaten ecosystem services^{7,8}.

It is notably hard to predict such critical transitions, because the state of the system may show little change before the tipping point is reached. Also, models of complex systems are usually not accurate enough to predict reliably where critical thresholds may occur. Interestingly, though, it now appears that certain generic symptoms may occur in a wide class of systems as they approach a critical point. At first sight, it may seem surprising that disparate phenomena such as the collapse of an overharvested population and ancient climatic transitions could be indicated by similar signals. However, as we will explain here, the dynamics of systems near a critical point have generic properties, regardless of differences in the details of each system⁹. Therefore, sharp transitions in a range of complex systems are in fact related. In models, critical thresholds for such transitions correspond to bifurcations¹⁰. Particularly relevant are ‘catastrophic bifurcations’ (see Box 1 for an example), where, once a threshold is exceeded, a positive feedback propels the system through a phase of directional change towards a contrasting state. Another important class of bifurcations are those that mark the transition from a stable equilibrium to a cyclic or chaotic attractor. Fundamental shifts that occur in systems when they pass bifurcations are collectively referred to as critical transitions¹¹.

We will first highlight the theoretical background of leading indicators that may occur in non-equilibrium dynamics before critical transitions, and illustrate how such indicators can perform in model generated time-series. Subsequently, we will review emerging empirical work on different systems and discuss prospects and challenges.

Theory

Critical slowing down and its symptoms. The most important clues that have been suggested as indicators of whether a system is getting close to a critical threshold are related to a phenomenon known in dynamical systems theory as ‘critical slowing down’¹². Although critical slowing down occurs for a range of bifurcations, we will focus on the fold catastrophe (Box 1) as a starting point. Inappropriate use of this classical model caused some controversy in the past¹³, but it is now

considered to capture the essence of shifts at tipping points in a wide range of natural systems ranging from cell signalling pathways¹⁴ to ecosystems^{7,15} and the climate⁶. At fold bifurcation points (F_1 and F_2 , Box 1), the dominant eigenvalue characterizing the rates of change around the equilibrium becomes zero. This implies that as the system approaches such critical points, it becomes increasingly slow in recovering from small perturbations (Fig. 1). It can be proven that this phenomenon will occur in any continuous model approaching a fold bifurcation¹². Moreover, analysis of various models shows that such slowing down typically starts far from the bifurcation point, and that recovery rates decrease smoothly to zero as the critical point is approached¹⁶. Box 2 describes a simple example illustrating this.

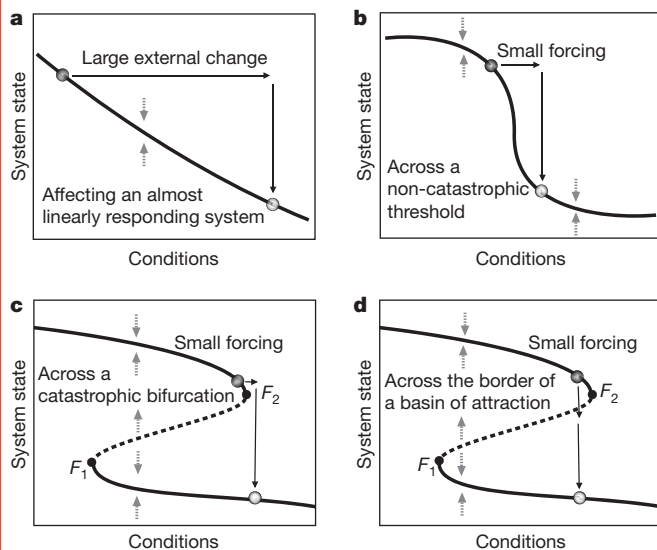
The most straightforward implication of critical slowing down is that the recovery rate after small experimental perturbation can be used as an indicator of how close a system is to a bifurcation point¹⁶. Because it is the rate of change close to the equilibrium that matters, such perturbations may be very small, posing no risk of driving the system over the threshold. Also, models indicate that in spatially extensive systems at risk of systemic collapse, small-scale experimental probing may suffice to test the vicinity of the threshold for such a large-scale transition. For instance, it has been shown that recovery times after local perturbation increase in models of fragmented populations approaching a threshold for global extinction¹⁷.

For most natural systems, it would be impractical or impossible to monitor them by systematically testing recovery rates. However, almost all real systems are permanently subject to natural perturbations. It can be shown that as a bifurcation is approached in such a system, certain characteristic changes in the pattern of fluctuations are expected to occur. One important prediction is that the slowing down should lead to an increase in autocorrelation in the resulting pattern of fluctuations¹⁸ (Fig. 1). This can be shown mathematically (Box 3), but it is also intuitively simple to understand. Because slowing down causes the intrinsic rates of change in the system to decrease, the state of the system at any given moment becomes more and more like its past state. The resulting increase in ‘memory’ of the system can be measured in a variety of ways from the frequency spectrum of the system^{19,20}. The simplest approach is to look at lag-1 autocorrelation^{21,22}, which can be directly interpreted as slowness of recovery in such natural perturbation regimes^{16,18}. Analyses of simulation models exposed to stochastic forcing confirm that if the system is driven gradually closer to a catastrophic bifurcation, there is a marked increase in autocorrelation that builds up long before the

¹Department of Environmental Sciences, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands. ²Integrative Ecology Group, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Calle Américo Vespucio s/n, E-41092 Sevilla, Spain. ³Department of Economics, 1180 Observatory Drive, ⁴Center for Limnology, 680 North Park Street, University of Wisconsin, Madison, Wisconsin 53706, USA. ⁵Max Planck Institute for Meteorology, Bundesstraße 55, 20146 Hamburg, Germany. ⁶Potsdam Institute for Climate Impact Research, PO Box 601203, D-14412 Potsdam, Germany. ⁷Utrecht University, Faculty of Geosciences, Department Environmental Sciences, Copernicus Institute, PO Box 80115, 3508 TC Utrecht, The Netherlands. ⁸Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

Box 1 | Critical transitions in the fold catastrophe model

The equilibrium state of a system can respond in different ways to changes in conditions such as exploitation pressure or temperature rise (Box 1 Figure a, b, c). If the equilibrium curve is folded backwards (Box 1 Figure c, d), three equilibria can exist for a given condition. The grey dotted arrows in the plots indicate the direction in which the system moves if it is not in equilibrium (that is, not on the curve). It can be seen from these arrows that all curves represent stable equilibria, except for the dashed middle section in Box 1 Figure c, d. If the system is driven slightly away from this part of the curve, it will move further away instead of returning. Hence, equilibria on this part of the curve are unstable and represent the border between the basins of attraction of the two alternative stable states on the upper and lower branches. If the system is very close to a fold bifurcation point (for example point F_1 or point F_2), a tiny change in the condition may cause a large shift in the lower branch (Box 1 Figure c). Also, close to such a bifurcation a small perturbation can drive the system across the boundary between the attraction basins (Box 1 Figure d). Thus, those bifurcation points are tipping points at which a tiny perturbation can produce a large transition. Small perturbations can also cause large changes in the absence of true bifurcations, provided that the system is very sensitive in a certain range of conditions (Box 1 Figure b). Finally, a shift in system state may simply be caused by a sudden large external force (Box 1 Figure a). Early-warning signals tend to arise as systems approach a bifurcation point such as in Box 1 Figure c, d, and also if systems approach a non-catastrophic threshold such as the one shown in Box 1 Figure b.



critical transition occurs (Fig. 2d). This is true not only for simple models²², but also for highly elaborate and relatively realistic models of spatially complex systems²³.

Increased variance in the pattern of fluctuations is another possible consequence of critical slowing down as a critical transition is approached²⁴ (Fig. 1). Again, this can be formally shown²⁵ (Box 3), as well as intuitively understood: as the eigenvalue approaches zero, the impacts of shocks do not decay, and their accumulating effect increases the variance of the state variable. In principle, critical slowing down could reduce the ability of the system to track the fluctuations, and thereby produce an opposite effect on the variance^{26,27}. However, analyses of models show that an increase in the variance usually arises and may be detected well before a critical transition occurs²⁴ (Fig. 2).

In summary, the phenomenon of critical slowing down leads to three possible early-warning signals in the dynamics of a system approaching a bifurcation: slower recovery from perturbations, increased autocorrelation and increased variance.

Skewness and flickering before transitions. In addition to autocorrelation and variance, the asymmetry of fluctuations may increase

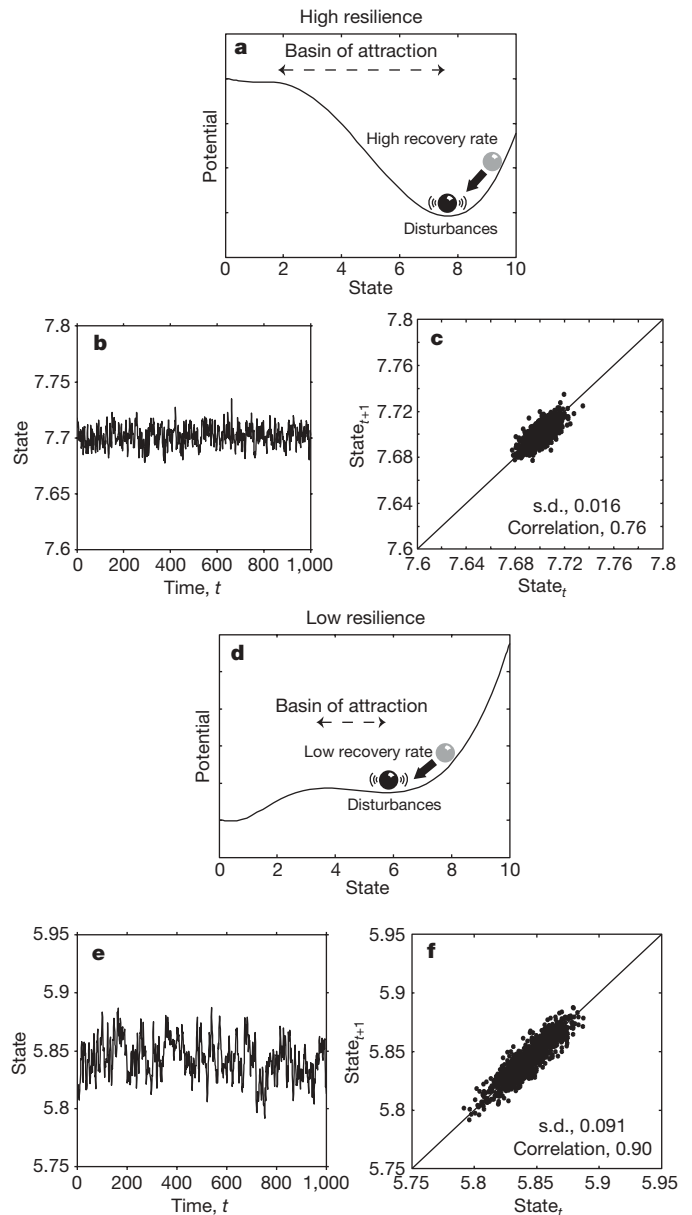


Figure 1 | Some characteristic changes in non-equilibrium dynamics as a system approaches a catastrophic bifurcation (such as F_1 or F_2 , Box 1).

a, b, c, Far from the bifurcation point (**a**), resilience is large in two respects: the basin of attraction is large and the rate of recovery from perturbations is relatively high. If such a system is stochastically forced, the resulting dynamics are characterized by low correlation between the states at subsequent time intervals (**b, c**). **d–f,** When the system is closer to the transition point (**d**), resilience decreases in two senses: the basin of attraction shrinks and the rate of recovery from small perturbations is lower. As a consequence of this slowing down, the system has a longer memory for perturbations, and its dynamics in a stochastic environment are characterized by a larger s.d. and a stronger correlation between subsequent states (**e, f**). Plots produced from a stochastically forced differential equation¹⁵ representing a harvested population: $dX/dt = X(1 - X/K) - c(X^2/(X^2 + 1))$, where X is population density, K is the carrying capacity (set to 10) and c is the maximum harvest rate (set to 1 for high resilience and 2.6 for low resilience).

before a catastrophic bifurcation²⁸. This does not result from critical slowing down. Instead, the explanation is that in catastrophic bifurcations such as fold bifurcations (Box 1), an unstable equilibrium that marks the border of the basin of attraction approaches the attractor from one side (Box 1). In the vicinity of this unstable point, rates of change are lower (reflected in a less steep slope in the stability landscapes). As a result, the system will tend to stay in the vicinity of

Box 2 | Critical slowing down: an example

To see why the rate of recovery rate after a small perturbation will be reduced, and will approach zero when a system moves towards a catastrophic bifurcation point, consider the following simple dynamical system, where γ is a positive scaling factor and a and b are parameters:

$$\frac{dx}{dt} = \gamma(x-a)(x-b) \quad (1)$$

It can easily be seen that this model has two equilibria, $\bar{x}_1 = a$ and $\bar{x}_2 = b$, of which one is stable and the other is unstable. If the value of parameter a equals that of b , the equilibria collide and exchange stability (in a transcritical bifurcation). Assuming that \bar{x}_1 is the stable equilibrium, we can now study what happens if the state of the equilibrium is perturbed slightly ($x = \bar{x}_1 + \varepsilon$):

$$\frac{d(\bar{x}_1 + \varepsilon)}{dt} = f(\bar{x}_1 + \varepsilon)$$

Here $f(x)$ is the right hand side of equation (1). Linearizing this equation using a first-order Taylor expansion yields

$$\frac{d(\bar{x}_1 + \varepsilon)}{dt} = f(\bar{x}_1 + \varepsilon) \approx f(\bar{x}_1) + \left. \frac{\partial f}{\partial x} \right|_{\bar{x}_1} \varepsilon$$

which simplifies to

$$f(\bar{x}_1) + \frac{d\varepsilon}{dt} = f(\bar{x}_1) + \left. \frac{\partial f}{\partial x} \right|_{\bar{x}_1} \varepsilon \Rightarrow \frac{d\varepsilon}{dt} = \lambda_1 \varepsilon \quad (2)$$

With eigenvalues λ_1 and λ_2 in this case, we have

$$\lambda_1 = \left. \frac{\partial f}{\partial x} \right|_a = -\gamma(b-a) \quad (3)$$

and, for the other equilibrium

$$\lambda_2 = \left. \frac{\partial f}{\partial x} \right|_b = \gamma(b-a) \quad (4)$$

If $b > a$ then the first equilibrium has a negative eigenvalue, λ_1 , and is thus stable (as the perturbation goes exponentially to zero; see equation (2)). It is easy to see from equations (3) and (4) that at the bifurcation ($b = a$) the recovery rates λ_1 and λ_2 are both zero and perturbations will not recover. Farther away from the bifurcation, the recovery rate in this model is linearly dependent on the size of the basin of attraction ($b - a$). For more realistic models, this is not necessarily true but the relation is still monotonic and is often nearly linear¹⁶.

the unstable point relatively longer than it would on the opposite side of the stable equilibrium. The skewness of the distribution of states is expected to increase not only if the system approaches a catastrophic bifurcation, but also if the system is driven closer to the basin boundary by an increasing amplitude of perturbation²⁸.

Another phenomenon that can be seen in the vicinity of a catastrophic bifurcation point is flickering. This happens if stochastic forcing is strong enough to move the system back and forth between the basins of attraction of two alternative attractors as the system enters the bistable region before the bifurcation^{26,29}. Such behaviour is also considered an early warning, because the system may shift permanently to the alternative state if the underlying slow change in conditions persists, moving it eventually to a situation with only one stable state. Flickering has been shown in models of lake eutrophication²⁴ and trophic cascades³⁰, for instance. Also, as discussed below, data suggest that certain climatic shifts and epileptic seizures may be presaged by flickering. Statistically, flickering can be observed in the frequency distribution of states as increased variance and skewness as well as bimodality (reflecting the two alternative regimes)²⁴.

Indicators in cyclic and chaotic systems. The principles discussed so far apply to systems that may be stochastically forced but have an underlying attractor that corresponds to a stable point (for example the classic fold catastrophe illustrated in Box 1). Critical transitions in cyclic and chaotic systems are less well studied from the point of view

Box 3 | The relation between critical slowing down, increased autocorrelation and increased variance

Critical slowing down will tend to lead to an increase in the autocorrelation and variance of the fluctuations in a stochastically forced system approaching a bifurcation at a threshold value of a control parameter. The example described here illustrates why this is so. We assume that there is a repeated disturbance of the state variable after each period Δt (that is, additive noise). Between disturbances, the return to equilibrium is approximately exponential with a certain recovery speed, λ . In a simple autoregressive model this can be described as follows:

$$x_{n+1} - \bar{x} = e^{\lambda \Delta t} (x_n - \bar{x}) + \sigma \varepsilon_n$$

$$y_{n+1} = e^{\lambda \Delta t} y_n + \sigma \varepsilon_n$$

Here y_n is the deviation of the state variable x from the equilibrium, ε_n is a random number from a standard normal distribution and σ is the standard deviation.

If λ and Δt are independent of y_n , this model can also be written as a first-order autoregressive (AR(1)) process:

$$y_{n+1} = \alpha y_n + \sigma \varepsilon_n$$

The autocorrelation $\alpha \equiv e^{\lambda \Delta t}$ is zero for white noise and close to one for red (autocorrelated) noise. The expectation of an AR(1) process $y_{n+1} = c + \alpha y_n + \sigma \varepsilon_n$ is¹⁸

$$E(y_{n+1}) = E(c) + \alpha E(y_n) + E(\sigma \varepsilon_n) \Rightarrow \mu = c + \alpha \mu + 0 \Rightarrow \mu = \frac{c}{1-\alpha}$$

For $c = 0$, the mean equals zero and the variance is found to be

$$\text{Var}(y_{n+1}) = E(y_n^2) - \mu^2 = \frac{\sigma^2}{1-\alpha^2}$$

Close to the critical point, the return speed to equilibrium decreases, implying that λ approaches zero and the autocorrelation α tends to one. Thus, the variance tends to infinity. These early-warning signals are the result of critical slowing down near the threshold value of the control parameter.

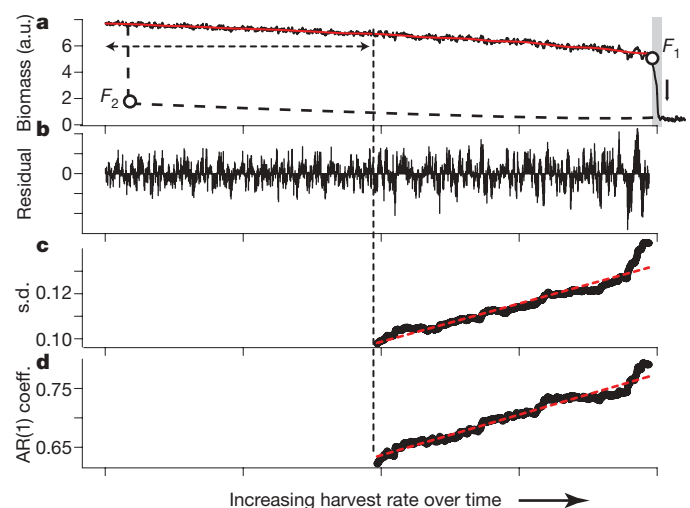


Figure 2 | Early warning signals for a critical transition in a time series generated by a model of a harvested population⁷⁷ driven slowly across a bifurcation. **a**, Biomass time series. **b**, **c**, **d**, Analysis of the filtered time series (**b**) shows that the catastrophic transition is preceded by an increase both in the amplitude of fluctuation, expressed as s.d. (**c**), and in slowness, estimated as the lag-1 autoregression (AR(1)) coefficient (**d**), as predicted from theory. The grey band in **a** identifies the transition phase. The horizontal dashed arrow shows the width of the moving window used to compute the indicators shown in **c** and **d**, and the red line is the trend used for filtering (see ref. 22 for the methods used). The dashed curve and the points F_1 and F_2 represent the equilibrium curve and bifurcation points as in Box 1 Figure c, d. a.u., arbitrary units.

of early-warning signals. Such transitions are associated with different classes of bifurcation¹⁰. First, there are the bifurcations that mark the transitions between stable, cyclic and chaotic regimes. An example is the Hopf bifurcation, which marks the transition from a stable system to an oscillatory system³¹. Like the fold bifurcation, this bifurcation is signalled by critical slowing down³²: close to the bifurcation, perturbations lead to long transient oscillations before the system settles to the stable state.

Another class of bifurcations are the non-local bifurcations¹⁰ that occur if intrinsic oscillations bring the system to the border of the basin of attraction of an alternative attractor. Such basin-boundary collisions³³ are not associated with particular properties of stable or unstable points that can be analytically defined. We know of no explicit work on early-warning signals for such transitions. Nonetheless, the dynamics may be expected to change in a characteristic way before basin-boundary collisions occur. For instance, oscillations may become 'stretched', as the system dwells longer in the vicinity of the basin boundary, where rates of change are slower³⁴, implying increased autocorrelation. Finally, there is the phenomenon of phase locking between coupled oscillators. Again, alternative attractors are often involved³⁵ and the corresponding bifurcations are associated with critical slowing down³⁶, suggesting the existence of early-warning signals. Indeed, rising variance and flickering occur before an epileptic seizure, a phenomenon associated to the phase locking of firing in neural cells (see below).

Spatial patterns as early-warning signals. In addition to early-warning signals in time series, there are particular spatial patterns that can arise before a critical transition. Many systems can be seen as consisting of numerous coupled units each of which tends to take a state similar to that of the units to which it is connected. For instance, it is well known that financial markets affect each other. Also, the attitudes of individuals towards certain issues is affected by the attitudes of their peers^{37,38}, and the persistence of species in habitat patches in a fragmented landscape depends on the presence of the same species in neighbouring patches from which recolonization can happen^{39,40}. In such systems, phase transitions may occur^{9,41} much as in ferromagnetic materials, where individual particles affect each others' spin. As gradual change in an external forcing factor (for example a magnetic field) drives the system closer to a transition, the distribution of the states of the units in such systems may change in characteristic ways. For instance, scale-invariant distributions of patch sizes occur close to a systemic transition, and there is a general tendency towards increased spatial coherence, measured as increased cross-correlation (or in oscillating units, resonance) among units before a critical event^{9,41}.

Certain classes of spatial system deviate from this general pattern and can have other, more specific, indicators of imminent transitions. For instance, in systems governed by local disturbance (for example grazers foraging locally on vegetation patches), scale-invariant power-law structures that are found for a large parameter range vanish as a critical transition is approached⁴². In systems that have self-organized regular patterns⁴³, critical transitions may be signalled by particular spatial configurations. For instance, models of desert vegetation show that as a critical transition to a barren state is neared, the vegetation becomes characterized by regular patterns because of a symmetry-breaking instability. These patterns change in a predictable way as the critical transition to the barren state is approached (Fig. 3), implying that this may be interpreted as early-warning signal for a catastrophic bifurcation⁴⁴.

In conclusion, when it comes to interpreting spatial patterns it is important to know which class of system is involved. Although broad classes have similar early-warning signals, there is no 'one-size-fits-all' spatial pattern announcing critical transitions.

Precursors of transitions in real systems

Most of the work on early-warning signals for critical transitions has so far been done using simple models, and empirical proof that critical slowing down occurs at bifurcations has been provided by controlled

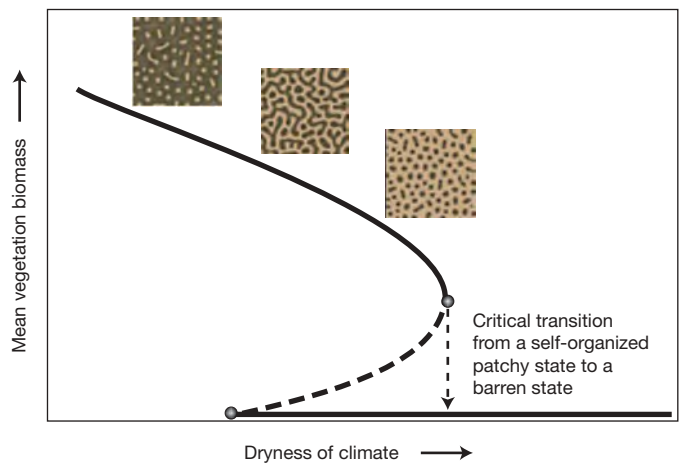


Figure 3 | Ecosystems may undergo a predictable sequence of self-organized spatial patterns as they approach a critical transition. We show the modelled response of semi-arid vegetation to increasing dryness of the climate. Solid lines represent mean equilibrium densities of vegetation. The insets are maps of the pattern: the dark colour represents vegetation and the light colour represents empty soil. As the bifurcation point for a critical transition into a barren state is approached, the nature of pattern changes from maze-like to spots. Modified from ref. 44. Reprinted with permission from AAAS.

experiments with lasers⁴⁵ and neurons⁴⁶. The question therefore remains of whether highly complex real systems such as the climate or ecosystems will show the theoretically expected early-warning signals. Results from elaborate and relatively realistic climate models including spatial dynamics and chaotic elements²³ suggest that some signals might be robust in the sense that they arise despite high complexity and noisiness. Nonetheless, it is clearly more challenging to pick up early-warning signals in complex natural systems than in models. We now review some emerging results on the climate and ecosystems. Also, we highlight empirical successes in finding early-warning signals of transitions in systems for which we have a relatively poor understanding of the mechanisms that drive the dynamics, such as the human brain and financial markets.

Climate. Interest in the possibility of critical transitions in the Earth system has been sparked by records of past climate dynamics revealing occasional sharp transitions from one regime to another⁴⁷. For instance, about 34 Myr ago the Earth changed suddenly from the tropical state in which it had been for many millions of years to a colder state in which Antarctica was glaciated, a shift known as the greenhouse–icehouse transition^{48–50} (Fig. 4). Also, glacial cycles tend to end with an abrupt warming^{51,52}.

Uncertainty in reconstructing such dynamics remains considerable, and it is even more difficult to unveil the underlying mechanisms. Nonetheless, the sharpness of the shifts and the existence of positive feedbacks that, if strong enough, could cause self-propelling change have led to the suggestion that these and other examples of rapid climate change could be explained as critical transitions^{6,47,53}. Therefore, the reconstructed climate dynamics before such transitions are an obvious place to look for early-warning signals. In a recent analysis, a significant increase in autocorrelation was found in each of eight examples of abrupt climate change analysed²² (Fig. 4).

Another recent study suggests that flickering preceded the abrupt end of the Younger Dryas cold period⁵⁴. Although the first part of this cold episode was quite stable, rapid alternations between a cold mode and a warm mode characterized the later part, and the episode eventually ending in a sharp shift to the relatively warm and stable conditions of the Holocene epoch⁵⁵. After examination of longer timescales, it has been suggested that the increasing Pleistocene climate variability may be interpreted as a signal that the near geological future might bring a transition from glacial–interglacial oscillations to a stable state characterized by permanent mid-latitude Northern Hemisphere glaciation⁵⁶.

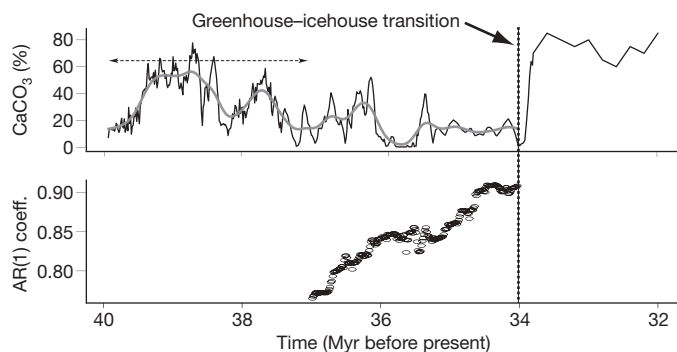


Figure 4 | Critical slowing down indicated by an increase in lag-1 autocorrelation in climate dynamics. We show the period preceding the transition from a greenhouse state to an icehouse state on the Earth 34 Myr ago. The trends in the CaCO_3 concentration time series removed by filtering before computing autocorrelation (AR(1) coefficient) are represented by the grey line. The horizontal dashed arrow shows the width of the moving window used to compute the autocorrelation. Modified from ref. 22.

Ecosystems. In ecology, critical transitions have become a major focus of research. The existence of alternative attractors has been demonstrated experimentally in lakes⁵⁷, and a large body of work now suggests that alternative stable states separated by critical thresholds also occur in ecosystems ranging from rangelands to marine systems^{7,58}. Work on early-warning signals in this field is just emerging. As mentioned earlier, in dry regions self-organization can lead to particular spatial patterns under some conditions. Here the complete loss of vegetation is an important transition, as recovery from the barren state may require more rain than is needed to preserve the last patches. There is good evidence to support the idea that a regular pattern characterized by spots of vegetation signals the proximity of a threshold to such catastrophic desertification⁴⁴. Other studies show how, in line with model predictions, vegetation-patch size distributions lose their scale-free structures and become characterized by truncated power laws as a transition to a barren state is approached⁴².

Early-warning signals are also being found for destabilization of exploited fish stocks. It has been shown that harvesting tends to lead to increased fluctuations in fish populations⁵⁹. This increase in variance is most likely due to increased intrinsic growth rates in the resulting populations, as older age classes are preferentially harvested and the younger fish have higher overall intrinsic rates of change⁶⁰. Such higher growth rates lead to increased nonlinearity as they drive populations towards the critical transition from a stable to a cyclic or chaotic regime, as mentioned earlier.

Asthma attacks and epileptic seizures. Abrupt transitions in physiology include epileptic seizures and asthma attacks. In the case of asthma, it has been shown that human lungs can display a self-organized pattern of bronchoconstriction that might be the prelude to dangerous respiratory failure, and which resembles the pattern formation in collapsing desert vegetation¹.

Epileptic seizures happen when neighbouring neural cells all start firing in synchrony. Predicting such seizures far in advance remains very difficult⁶¹. However, before the seizure becomes noticeable several characteristic changes in neural activity can occur. For instance, minutes before an epileptic seizure, variance in the electrical signal recorded by electroencephalography may increase³ (Fig. 5). More subtle changes (reduced dimensionality of the signal) occur up to 25 min before epileptic seizures, reflecting a continuous increase in the degree of synchronicity (and thus correlation) between neural cells⁶². Also, hours before the seizure, mild energy bursts can occur in the brain followed by frequent symptomless seizures too small for the patients to notice². This resembles patterns of flickering in which smaller transient excursions to the vicinity of an alternative state precede the upcoming major shift.

Finance. The prediction of shifts in financial markets is a heavily researched area. In this field, the discovery of predictability quickly

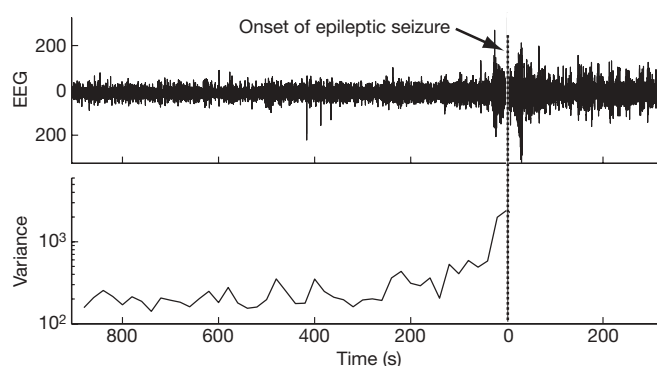


Figure 5 | Subtle changes in brain activity before an epileptic seizure may be used as an early warning signal. The epileptic seizure clinically detected at time 0 is announced minutes earlier in an electroencephalography (EEG) time series by an increase in variance. Adapted by permission from Macmillan Publishers Ltd: Nature Medicine (ref. 3), copyright 2003.

leads to its elimination, as profit can be made from it, thereby annihilating the pattern. As a result, although there is always some predictability that can be exploited by specialists^{63,64}, overall financial markets are notoriously difficult to predict⁶⁵. Nonetheless, many papers in the financial literature show that market dynamics may contain information presaging major events^{66–68}. For instance, some events are preceded by measures of increased trade volatility (for example the spread between the value of put options and the value of call options^{66,67}), but a ‘volatility calm’ before the abrupt change can also happen^{66,69}. A prominent volatility-based early-warning signal in financial markets is the VIX, or ‘fear’ index^{70,71}. There is also evidence of systematic relationships in variance and first-order autocorrelation⁷², although lead–lag relationships tend to be erratic. Finally, increased spatial coherence may be an early warning of major transitions. There is evidence that correlation increases across returns to different stocks in a falling market and patterns embodied in options prices may serve as a type of early-warning indicator⁷³.

Outlook

As our overview shows, similar early-warning signals can appear in widely different systems: flickering may occur before epileptic seizures, the end of a glacial period and in lakes before they shift to a turbid state; self-organized patterns can signal an imminent transition in desert vegetation and in asthma; increased autocorrelation may indicate critical slowing down before all kinds of climatic transitions and in ecosystems; and increased variance of fluctuation may be a leading indicator of an epileptic seizure or instability in an exploited fish stock. Some of these complex systems are better understood than others. However, turning the reasoning around, it could be argued that the generic character of some early-warning signals suggests that these transitions may be somehow related to bifurcations, where universal laws of dynamical systems govern the pattern.

The theoretical basis of the work on early-warning signals in simple models is quite strong, and the first results from more elaborate models suggest that similar signals may arise in highly complex systems²³. Nonetheless, more work is needed to find out how robust these signals are in situations in which spatial complexity, chaos and stochastic perturbations govern the dynamics. Also, detection of the patterns in real data is challenging and may lead to false positive results as well as false negatives. False negatives are situations in which a sudden transition occurred but no early-warning signals could be detected in the behaviour before the shift. This can happen for different reasons. One possibility is that the sudden shift in the system was not preceded by a gradual approach to a threshold. For instance, it may have remained at the same distance from the bifurcation point, but been driven to another stable state by a rare extreme event. Also, a shift that is simply due to a fast and permanent change of external conditions (Box 1 Figure a) cannot be detected from early-warning signals. A

second class of false negatives may arise from the statistical difficulty of picking up the early-warning signal. For instance, the detection of increased autocorrelation may require long time series⁷⁴. A third difficulty arises if the external regime of perturbations changes over time. This may distort or counteract the expected signals. False positives occur if a supposed early-warning signal is not the result of approaching a bifurcation. This may happen by chance or may result from a confounding trend within the system or in the external regime of perturbations.

Importantly, most of the indicators we have identified signal a wide class of impending transitions in complex systems. The same signals may even occur, albeit in a less pronounced way, as the system approaches a threshold that is not related to catastrophic bifurcations (Box 1 Figure b)²⁷. This has been shown for critical slowing down¹⁶, and may also be true for autocorrelation and variance. Nonetheless, such non-catastrophic thresholds are related to the more spectacular catastrophic ones (Box 1), and systems may in fact move from one type of threshold to another. In conclusion, most early-warning signals are indicators of proximity to a broad class of thresholds, where small forces can cause major changes in the state of a complex system.

The idea that critical transitions across a range of systems may be related in the sense that they can be described by similar equations, implying similar possible bifurcations and early-warning signals, implies an exciting opportunity for connecting work across disciplines. However, there are many challenges to be overcome. For instance, filtering techniques for time series⁷⁵ are necessary to increase the sensitivity of indicators while preventing false positives²², but results depend on parameter choices in filtering^{22,23}. Therefore, it would be useful to build a set of reliable statistical procedures to test whether an increase in autocorrelation, for example, is significant. We note also that most of the signals we have discussed should still be interpreted in a relative sense. For instance, although autocorrelation is predicted to approach unity at a fold bifurcation, measurement noise will tend to reduce correlations. Also, perturbations will often trigger a transition well before a bifurcation point is reached. Thus, although a trend in the indicators may serve as a warning, the actual moment of a transition remains difficult to predict. A key issue when it comes to practical application is the question of whether a signal can be detected sufficiently early for action to be taken to prevent a transition or to prepare for one²⁵. Understanding spatial early-warning signals better might be particularly useful in this respect, as a spatial pattern contains much more information than does a single point in a time series, in principle allowing shorter lead times⁷⁶.

In any case, generic early-warning signals will remain only one of the tools we have for predicting critical transitions. In systems in which we can observe transitions repeatedly, such as lakes, rangelands or fields such as physiology, we may empirically discover where the thresholds are. Nonetheless, some extremely important systems, such as the climate or ocean circulation, are singular and afford us limited opportunity to learn by studying many similar transitions. Also, we are far from being able to develop accurate models to predict thresholds in most complex systems, ranging from cells to organisms, ecosystems or the climate. We simply do not understand all the relevant mechanisms and feedbacks sufficiently well in most cases. The generic character of the early-warning signals we have discussed here is reason for optimism, as they occur largely independently of the precise mechanism involved. Thus, if we have reasons to suspect the possibility of a critical transition, early-warning signals may be a significant step forwards when it comes to judging whether the probability of such an event is increasing.

1. Venegas, J. G. *et al.* Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* **434**, 777–782 (2005).
2. Litt, B. *et al.* Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* **30**, 51–64 (2001).
3. McSharry, P. E., Smith, L. A. & Tarassenko, L. Prediction of epileptic seizures: are nonlinear methods relevant? *Nature Med.* **9**, 241–242 (2003).

4. Kambhu, J., Weidman, S. & Krishnan, N. *New Directions for Understanding Systemic Risk: A Report on a Conference Cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences* (The National Academies Press, 2007).
5. May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451**, 893–895 (2008).
6. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
7. Scheffer, M. *et al.* Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
8. Millennium Ecosystem Assessment. *Ecosystems and Human Well-being: Synthesis Report* (Island, 2005).
9. Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (Freeman, 1991).
10. Kuznetsov, Y. A. *Elements of Applied Bifurcation Theory* (Springer, 1995).
11. Scheffer, M. *Critical Transitions in Nature and Society* (Princeton Univ. Press, 2009).
- This book is an introduction to the theory of critical transitions with examples from ecosystems, climate dynamics and social systems.
12. Wissel, C. A universal law of the characteristic return time near thresholds. *Oecologia* **65**, 101–107 (1984).
13. Zahler, R. S. & Sussmann, H. J. Claims and accomplishments of applied catastrophe theory. *Nature* **269**, 759–763 (1977).
14. Bagowski, C. P. & Ferrell, J. E. Bistability in the JNK cascade. *Curr. Biol.* **11**, 1176–1182 (2001).
15. May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
16. Van Nes, E. H. & Scheffer, M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am. Nat.* **169**, 738–747 (2007).
- This paper shows that recovery rate after disturbances is a good indicator of the proximity of a critical threshold.
17. Ovaskainen, O. & Hanski, I. Transient dynamics in metapopulation response to perturbation. *Theor. Popul. Biol.* **61**, 285–295 (2002).
18. Ives, A. R. Measuring resilience in stochastic systems. *Ecol. Monogr.* **65**, 217–233 (1995).
19. Kleinen, T., Held, H. & Petschel-Held, G. The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation. *Ocean Dyn.* **53**, 53–63 (2003).
20. Livina, V. N. & Lenton, T. M. A modified method for detecting incipient bifurcations in a dynamical system. *Geophys. Res. Lett.* **34**, L03712 (2007).
21. Held, H. & Kleinen, T. Detection of climate system bifurcations by degenerate fingerprinting. *Geophys. Res. Lett.* **31**, L23207 (2004).
- This paper shows that autocorrelation increases in the vicinity of a bifurcation in a model of the thermohaline circulation.
22. Dakos, V. *et al.* Slowing down as an early warning signal for abrupt climate change. *Proc. Natl Acad. Sci. USA* **105**, 14308–14312 (2008).
- This paper shows that autocorrelation increased before eight well-known climate transitions in the past, and describes the methodology of data processing required to do such an analysis.
23. Lenton, T. M. *et al.* Using GENIE to study a tipping point in the climate system. *Phil. Trans. R. Soc. A* **367**, 871–884 (2009).
24. Carpenter, S. R. & Brock, W. A. Rising variance: a leading indicator of ecological transition. *Ecol. Lett.* **9**, 308–315 (2006).
- This paper shows that variance increases in the vicinity of a bifurcation in a lake model.
25. Biggs, R., Carpenter, S. R. & Brock, W. A. Turning back from the brink: detecting an impending regime shift in time to avert it. *Proc. Natl Acad. Sci. USA* **106**, 826–831 (2009).
26. Berglund, N. & Gentz, B. Metastability in simple climate models: pathwise analysis of slowly driven Langevin equations. *Stoch. Dyn.* **2**, 327–356 (2002).
27. Berglund, N. & Gentz, B. *Noise-Induced Phenomena in Slow-Fast Dynamical Systems - A Sample-Paths Approach* (Springer, 2006).
28. Guttal, V. & Jayaprakash, C. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecol. Lett.* **11**, 450–460 (2008).
29. Horsthemke, W. & Lefever, R. *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology* (Springer, 1984).
30. Carpenter, S. R. *et al.* Leading indicators of trophic cascades. *Ecol. Lett.* **11**, 128–138 (2008).
31. Strogatz, S. H. *Nonlinear Dynamics and Chaos - With Applications to Physics, Biology, Chemistry and Engineering* 1st edn, 248–254 (Addison-Wesley, 1994).
32. Chisholm, R. A. & Filotas, E. Critical slowing down as an indicator of transitions in two-species models. *J. Theor. Biol.* **257**, 142–149 (2009).
33. Vandermeer, J. & Yodzis, P. Basin boundary collision as a model of discontinuous change in ecosystems. *Ecology* **80**, 1817–1827 (1999).
34. Rinaldi, S. & Scheffer, M. Geometric analysis of ecological models with slow and fast processes. *Ecosystems* **3**, 507–521 (2000).
35. Vandermeer, J., Stone, L. & Blasius, B. Categories of chaos and fractal basin boundaries in forced predator-prey models. *Chaos Solitons Fractals* **12**, 265–276 (2001).
36. Leung, H. K. Bifurcation of synchronization as a nonequilibrium phase transition. *Physica A* **281**, 311–317 (2000).
37. Scheffer, M., Westley, F. & Brock, W. Slow response of societies to new problems: causes and costs. *Ecosystems* **6**, 493–502 (2003).
38. Holyoak, J. A., Kacperski, K. & Schweitzer, F. Social impact models of opinion dynamics. *Annu. Rev. Comput. Phys.* **9**, 253–273 (2002).

39. Bascompte, J. & Solé, R. V. Habitat fragmentation and extinction thresholds in spatially explicit models. *J. Anim. Ecol.* **65**, 465–473 (1996).
40. Hanski, I. Metapopulation dynamics. *Nature* **396**, 41–49 (1998).
41. Solé, R. V. *et al.* Phase transitions and complex systems. *Complexity* **1**, 13–26 (1996).
42. Kefi, S. *et al.* Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. *Nature* **449**, 213–217 (2007).
43. Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B* **237**, 37–72 (1952).
44. Rietkerk, M., Dekker, S. C., de Ruiter, P. C. & van de Koppel, J. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* **305**, 1926–1929 (2004). **This paper discusses how self-organized vegetation patterns may indicate the vicinity of a catastrophic bifurcation in which all vegetation is lost.**
45. Tredicce, J. R. *et al.* Critical slowing down at a bifurcation. *Am. J. Phys.* **72**, 799–809 (2004).
46. Matsumoto, G. & Kunisawa, T. Critical slowing-down near the transition region from the resting to time-ordered states in squid giant axons. *J. Phys. Soc. Jpn* **44**, 1047–1048 (1978).
47. Alley, R. B. *et al.* Abrupt climate change. *Science* **299**, 2005–2010 (2003).
48. Tripati, A., Backman, J., Elderfield, H. & Ferretti, P. Eocene bipolar glaciation associated with global carbon cycle changes. *Nature* **436**, 341–346 (2005).
49. Kump, L. R. Foreshadowing the glacial era. *Nature* **436**, 333–334 (2005).
50. Liu, Z. *et al.* Global cooling during the Eocene-Oligocene climate transition. *Science* **323**, 1187–1190 (2009).
51. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
52. Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
53. Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
54. Bakke, J. *et al.* Rapid oceanic and atmospheric changes during the Younger Dryas cold period. *Nature Geosci.* **2**, 202–205 (2009).
55. Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415**, 863–869 (2002).
56. Crowley, T. J. & Hyde, W. T. Transient nature of late Pleistocene climate variability. *Nature* **456**, 226–230 (2008).
57. Scheffer, M. & van Nes, E. H. Shallow lakes theory revisited: various alternative regimes driven by climate, nutrients, depth and lake size. *Hydrobiologia* **584**, 455–466 (2007).
58. Scheffer, M. & Carpenter, S. R. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol. Evol.* **18**, 648–656 (2003).
59. Hsieh, C. H. *et al.* Fishing elevates variability in the abundance of exploited species. *Nature* **443**, 859–862 (2006).
60. Anderson, C. N. K. *et al.* Why fishing magnifies fluctuations in fish abundance. *Nature* **452**, 835–839 (2008). **This paper shows that increased variance in exploited fish stocks may arise from higher growth rates leading to stronger nonlinearity.**
61. Mormann, F., Andrzejak, R. G., Elger, C. E. & Lehnertz, K. Seizure prediction: the long and winding road. *Brain* **130**, 314–333 (2007).
62. Elger, C. E. & Lehnertz, K. Seizure prediction by non-linear time series analysis of brain electrical activity. *Eur. J. Neurosci.* **10**, 786–789 (1998).
63. Brock, W., Lakonishok, J. & Lebaron, B. Simple technical trading rules and the stochastic properties of stock returns. *J. Finance* **47**, 1731–1764 (1992).
64. Lo, A. W., Mamaysky, H. & Wang, J. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *J. Finance* **55**, 1705–1765 (2000).
65. Lebaron, B. The stability of moving average technical trading rules on the Dow Jones Index. *Deriv. Use Trad. Regul.* **5**, 324–338 (2000).
66. Bates, D. S. The crash of 87 - was it expected? The evidence from options markets. *J. Finance* **46**, 1009–1044 (1991).
67. Bates, D. S. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Rev. Financ. Stud.* **9**, 69–107 (1996).
68. Hens, T. & Schenk-Hoppe, K. R. (eds) *Handbook of Financial Markets: Dynamics and Evolution* (North-Holland, 2009).
69. Arvedlund, E. Calm before the storm? Low volatility often precedes market downturn. *Barron's* (January 28, 2002).
70. Whaley, R. E. Derivatives on market volatility: hedging tools long overdue. *J. Deriv.* **1**, 71–84 (1993).
71. Whaley, R. E. The investor fear gauge. *J. Portfol. Manage.* **26**, 12–17 (2000).
72. Lebaron, B. Some relations between volatility and serial correlations in stock-market returns. *J. Bus.* **65**, 199–219 (1992).
73. Hong, H. & Stein, J. C. Differences of opinion, short-sales constraints, and market crashes. *Rev. Financ. Stud.* **16**, 487–525 (2003).
74. Bence, J. R. Analysis of short-time series - correcting for autocorrelation. *Ecology* **76**, 628–639 (1995).
75. Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. *Time Series Analysis: Forecasting and Control* (Prentice-Hall, 1994).
76. Guttal, V. & Jayaprakash, C. Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems. *Theor. Ecol.* **11**, 450–460 (2008).
77. Van Nes, E. H. & Scheffer, M. Implications of spatial heterogeneity for regime shifts in ecosystems. *Ecology* **86**, 1797–1807 (2005).

Acknowledgements Our work was supported by the Institute Para Limes and by the South American Institute for Resilience and Sustainability Studies. The work of S.R.C. is also supported by the US National Science Foundation. The research of M.R. and V.D. is supported by the Netherlands Organisation of Scientific Research, Earth and Life Sciences, in the case of M.R. through a Vidi grant. The work of J.B. is supported by the European Heads of Research Councils, the European Science Foundation and the European Commission Sixth Framework Programme through a European Young Investigator Award. G.S. was supported by Deutsche Bank Jameson Complexity Studies Fund, and by an NSF/NOAA CAMEO Grant NA08OAR4320894.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to M.S. (marten.scheffer@wur.nl).

Scaling of BMP gradients in *Xenopus* embryos

Arising from: D. Ben-Zvi *et al.* *Nature* **453**, 1205–1211 (2008)

Metazoan organisms can ‘scale’, that is, maintain similar proportions regardless of size. Ben-Zvi *et al.*¹ use experiments in *Xenopus* to support a quantitative model that explains morphological scaling as the result of scaling of a gradient of bone morphogenetic protein (BMP) signals. We believe that the evidence for scaling in *Xenopus* is misinterpreted, and that their model for embryonic patterning disagrees with prior data. The experiments they present supporting their model admit alternative interpretations.

The authors’ model (box 1 of ref. 1) is built around the (BMP inhibitor) Chordin-facilitated transport of two members of the BMP family of ligands, BMP (BMP2/4/7), the total amount of which is preserved, and ADMP, which is produced only in the dorsal organizer but concentrates ventrally and scales along the dorsal–ventral axis at blastula stage.

The paper is based on the assumption that dorsal embryonic halves produce well-proportioned (scaled) tadpoles, which is in contrast to existing data. Kageura *et al.*² show that removing ventral cells from the eight-cell blastula (series 15, 17) results in normal heads attached to a small body. This is in accordance with standard fate maps that assign most of what is conventionally called ventral in the blastula to posterior tissue in the tadpole³. It is right–left half-embryos that will reproduce correctly proportioned half-size tadpoles². The quoted paper by Cooke⁴ examined only mesoderm patterning in transverse sections of tailbud embryos. Therefore, dorsal half-embryos do not scale in the sense defined by the authors. We henceforth focus on molecular evidence contradicting the presented theory construed as a model for embryonic patterning.

In the frog a twofold change in morphogen levels can elicit different cell fates⁵. Because ligands cannot be directly measured, nuclear Smad1/5/8 transcription factor is the best measure for total BMP signalling⁶. Experiments in frog⁷ and fish⁸ show at most a fourfold variation versus the 10² to 10⁴ range required for scaling in the model.

The model requires that total BMP activity derives predominantly from ADMP, yet BMP depletion (figure 2H in ref. 9) has a stronger phenotype than ADMP-depletion (figure 2H in ref. 10), resulting in embryos with disproportionately large heads^{8,9} similar to dorsal half-embryos (figure 3C in ref. 10). BMP4 injection significantly ventralizes the embryo (figure 1H in ref. 9), yet the model does not constrain the total amount of BMP or its initial location, because ‘shuttling’ actively concentrates it on the ventral side.

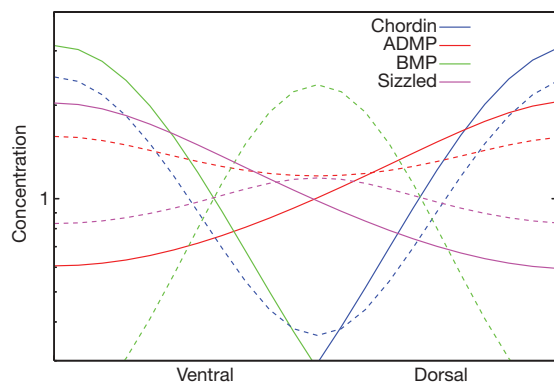


Figure 1 | Morphogen profiles after axes duplication (dashed lines) compared with the wild-type embryo (solid lines) from a reaction diffusion model¹⁴. The gene network is adapted from figure 3L in ref. 10 and involves self-sustained ventral and dorsal centres. The concentration of the ventral BMP marker in the lateral region between the duplicated axes is clearly reproduced. (The inhibitors, ADMP and Sizzled, diffuse rapidly, the other species slowly. Activation and repression are modelled as Hill functions and summarize more complex biochemistry. The equations and parameters that were solved to produce Fig. 1 are available from the authors.

Experiments in figure 3 of ref. 1 were performed to demonstrate Chordin-dependent shuttling. BMP4 is used instead of ADMP, and the protein distribution is shown in mid–late gastrula, although BMP must act before early gastrulation to affect dorsal–ventral patterning significantly¹¹. Labelled BMP4 is localized in endoderm and not ventral mesoderm as in the schematic of figure 3a of ref. 1. Other explanations for the localization of injected BMP4, such as secretion into the blastocoel cavity and ectopic uptake¹², need to be addressed. The Chordin-depleted embryos used as controls still show movement of injected BMP4, and the phenotype undercuts the larger message given that such embryos have well-defined axes¹³. The complete model that addresses these questions (supplementary information 6a–h of ref. 1) contains over 30 free constants to explain essentially qualitative data; a number so large as to render the predictions questionable.

Axis duplication experiments (figure 4 of ref. 1) are taken as evidence for shuttling: the authors assert that the (well-known) expression of ventral markers between the two axes is evidence for their mode of transport. However, there must be a maximum in BMP signalling between the two axes because it is suppressed in each. Reaction diffusion models^{14,15} show that ordinary diffusion, as opposed to facilitated diffusion through shuttling, can generate patterns consistent with the qualitative data presented (our Fig. 1).

In summary, we feel it is incorrect to appeal to qualitative data beyond the onset of gastrulation to support a model for blastula patterning, because other layers of regulation may intervene.

Paul Francois¹, Alin Vonica², Ali H. Brivanlou² & Eric D. Siggia¹

¹Center for Studies in Physics and Biology, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

e-mail: siggiae@rockefeller.edu

²Laboratory of Molecular Vertebrate Embryology, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

Received 14 November 2008; accepted 24 June 2009.

- Ben-Zvi, D., Shilo, B., Fainsod, A. & Barkai, N. Scaling of the BMP activation gradient in *Xenopus* embryos. *Nature* **453**, 1205–1211 (2008).
- Kageura, H. & Yamana, K. Pattern regulation in defect embryos of *Xenopus laevis*. *Dev. Biol.* **101**, 410–415 (1984).
- Gerhart, J. Changing the axis changes the perspective. *Dev. Dyn.* **225**, 380–383 (2002).
- Cooke, J. Scale of body pattern adjusts to available cell number in amphibian embryos. *Nature* **290**, 775–778 (1981).
- Kinoshita, T., Jullien, J. & Gurdon, J. B. Two-dimensional morphogen gradient in *Xenopus*: boundary formation and real-time transduction response. *Dev. Dyn.* **235**, 3189–3198 (2006).
- Faure, S., Lee, M. A., Keller, T., ten Dijke, P. & Whitman, M. Endogenous patterns of TGFβ superfamily signaling during early *Xenopus* development. *Development* **127**, 2917–2931 (2000).
- Schöhl, A. & Fagotto, F. Beta-catenin, MAPK and Smad signaling during early *Xenopus* development. *Development* **129**, 37–52 (2002).
- Tucker, J. A., Mintzer, K. A. & Mullins, M. C. The BMP signaling gradient patterns dorsoventral tissues in a temporally progressive manner along the anteroposterior axis. *Dev. Cell* **14**, 108–119 (2008).
- Reversade, B., Kuroda, H., Lee, H., Mays, A. & De Robertis, E. M. Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in *Xenopus* embryos. *Development* **132**, 3381–3392 (2005).
- Reversade, B. & De Robertis, E. Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell* **123**, 1147–1160 (2005).
- Marom, K., Levy, V., Pillemer, G. & Fainsod, A. Temporal analysis of the early BMP functions identifies distinct anti-organizer and mesoderm patterning phases. *Dev. Biol.* **282**, 442–454 (2005).
- Williams, P. H., Hagemann, A., Gonzalez-Gaitan, M. & Smith, J. C. Visualizing long-range movement of the morphogen Xnr2 in the *Xenopus* embryo. *Curr. Biol.* **14**, 1916–1923 (2004).
- Oelgeschläger, M., Kuroda, H., Reversade, B. & De Robertis, E. M. Chordin is required for the Spemann organizer transplantation phenomenon in *Xenopus* embryos. *Dev. Cell* **4**, 219–230 (2003).
- Meinhardt, H. Organizer and axes formation as a self-organizing process. *Int. J. Dev. Biol.* **45**, 177–188 (2001).
- Solnica-Krezel, L. Vertebrate development: taming the nodal waves. *Curr. Biol.* **13**, R7–R9 (2003).

doi:10.1038/nature08305

Reply to Francois et al.

Replying to: P. Francois et al. *Nature* 461, doi:10.1038/nature08305 (2009)

Francois *et al.*¹, commenting on our paper², argue that (1) scaling does not occur, (2) our model is inconsistent with existing experiments, and (3) our experiments are not conclusive. We disagree.

The ability of amphibian embryos to scale pattern with size is evident from the large variability in egg size, and is a pre-assumption of our study. In his 1938 monograph, H. Spemann reports his manipulation of newt embryos: "...when the two halves are completely separated, the dorsal half develops into a small embryo of normal proportions"³ (Fig. 1a). This experiment is cited in standard textbooks⁴ and reproduced in laboratories⁵ and teaching courses. Clearly, not all half-embryos develop normally, but this is not surprising given the likelihood of secondary damage. The key point is that surviving embryos maintain mesodermal ventral tissues (for example, blood and heart), expected to be lost in the absence of scaling.

The possibility that scaling results from over-growth was ruled out by J. Cooke, who demonstrated that proportionate assignment of trunk mesodermal cells to tissues is maintained in embryos in which as much as ~70% of the cells are removed⁶. The tissues examined

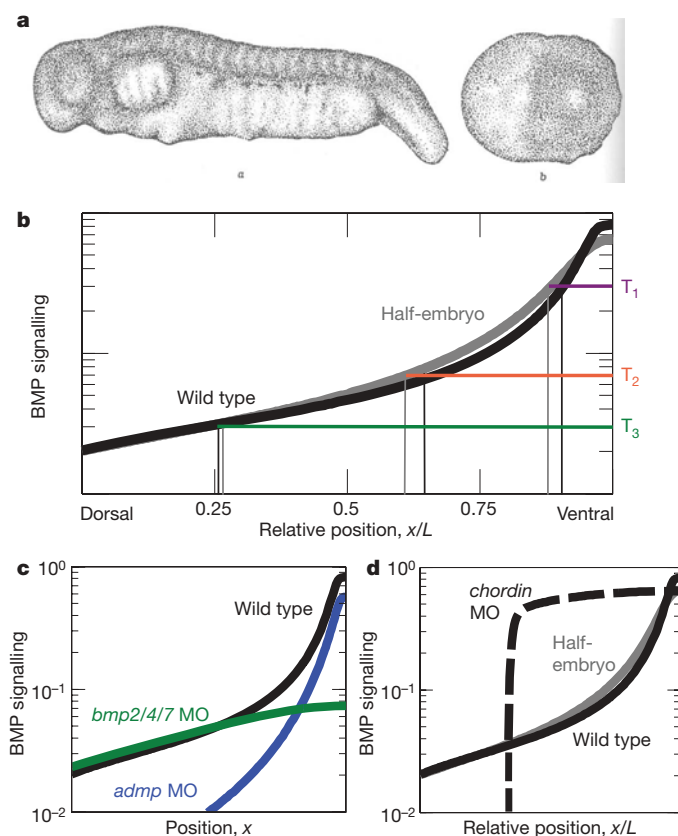


Figure 1 | Scaling pattern with size. **a**, A newt dorsal half-embryo (**a**, left) develops into a small embryo of normal proportions, in contrast to the ventral half (**b**, right). (Panel **a** is from ref. 3.) **b**, The BMP activation profile predicted by our model for wild-type (black line) and half-sized (grey line) embryos. L is the length of the dorsal-ventral axis and T_1 , T_2 and T_3 are arbitrary thresholds differing by less than tenfold. **c**, Numerical simulation of profiles in a wild-type embryo (black line), embryo depleted of Admp by *admp* morpholino (MO) (blue line) or depleted of Bmp2/4/7 (green line). **d**, Numerical simulation of profiles in wild-type (black line) and half-sized (grey line) embryos, and of an embryo depleted of Chordin (dashed black line, *chordin* morpholino). The same parameters were used in all figures (contact the authors for further information on parameters).

depend on bone morphogenetic protein (BMP) and thus provide readout of the early gradient.

Scaling in our model² is robust to the parameter choice, but the range of the gradient depends on parameters. We do not attempt to predict the *in vivo* parameters. The ~100-fold concentration difference shown (Fig. 1b) is comparable with the estimated range of the Dpp gradient in the significantly smaller *Drosophila* wing imaginal disc⁷. Quantitative measures of the pSmad1 profile in the opaque *Xenopus* embryo are still limited (see ref. 8 for example).

Our model is consistent with existing experiments. The severe phenotype of Bmp2/4/7 depletion (Fig. 1c) reflects the positive feedback on *bmp4* expression, which increases ventral Bmp4 levels. Similarly, the remaining polarity of Chordin-depleted embryos (Fig. 1d) is explained by the additional BMP inhibitors (for example, Noggin), which are structurally different from Chordin, and are not likely to be cleaved by Xlr or to mediate shuttling. Both properties were included in our original simulations and do not alter the robustness of the scaling mechanism.

Our model is robust to ligand production rate, but we do not expect this robustness to hold for arbitrarily high levels which can override available Chordin. Importantly, our conclusion that dorsally produced Admp accumulates ventrally was in fact tested: Bmp2/4/7-depleted embryos retain dorsal-ventral polarity⁹ and this polarity is abolished only when Admp is also depleted⁵.

Francois *et al.*¹ are concerned that separation of Bmp4-Myc from its site of injection may be due to ligand secretion into the blastocoel, and non-uniform uptake by ventral cells. However, separation was abolished when Chordin was depleted, with no apparent reason to assume that such secretion and uptake requires Chordin.

The novel point of our secondary-axis experiment is the removal of Admp from the secondary organizer. This eliminates an auxiliary source of a BMP ligand, which could contribute to the increase of BMP in mid-embryo. Francois *et al.* imply that a BMP profile that peaks in the centre of the embryo to a level similar to the normal 'lateral' level is sufficient to obtain *sizzled* expression. Clearly, this is not the case, because *sizzled* expression requires high BMP levels normally found in ventral positions.

Our model refers to the gastrulation stage, when BMP functions in dorsal-ventral patterning, and not the pre-gastrula stage, when BMP represses the organizer¹⁰. Francois *et al.* suggest that patterning can be explained by a different reaction-diffusion model. Their underlying assumptions¹¹, however, do not reflect the known network topology¹², and the resulting profiles are inconsistent with the system properties. Hence, we do not find their model to be a valid alternative.

Danny Ben-Zvi¹, Ben-Zion Shilo¹, Abraham Fainsod² & Naama Barkai^{1,3}

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: naama.barkai@weizmann.ac.il

²Department of Cellular Biochemistry and Human Genetics, Faculty of Medicine, Hebrew University, Jerusalem 91120, Israel.

³Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel.

1. Francois, P., Vonica, A., Brivanlou, A. H. & Siggia, E. D. Scaling of BMP gradients in *Xenopus* embryos. *Nature* doi:10.1038/nature08305 (this issue).
2. Ben-Zvi, D., Shilo, B. Z., Fainsod, A. & Barkai, N. Scaling of the BMP activation gradient in *Xenopus* embryos. *Nature* 453, 1205–1211 (2008).
3. Spemann, H. *Embryonic Development and Induction* 29–30 (Yale Univ. Press, 1938).
4. Gilbert, S. F. *Developmental Biology* 6th edn (Sinauer, 2000).
5. Reversade, B. & De Robertis, E. M. Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell* 123, 1147–1160 (2005).

6. Cooke, J. Scale of body pattern adjusts to available cell number in amphibian embryos. *Nature* **290**, 775–778 (1981).
7. Bollenbach, T. *et al.* Precision of the Dpp gradient. *Development* **135**, 1137–1146 (2008).
8. Schohl, A. & Fagotto, F. Beta-catenin, MAPK and Smad signaling during early *Xenopus* development. *Development*, **129**, 37–52 (2002).
9. Reversade, B., Kuroda, H., Lee, H., Mays, A. & De Robertis, E. M. Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in *Xenopus* embryos. *Development* **132**, 3381–3392 (2005).
10. Marom, K., Levy, V., Pillemer, G. & Fainsod, A. Temporal analysis of the early BMP functions identifies distinct anti-organizer and mesoderm patterning phases. *Dev. Biol.* **282**, 442–454 (2005).
11. Meinhardt, H. Organizer and axes formation as a self-organizing process. *Int. J. Dev. Biol.* **45**, 177–188 (2001).
12. Eivers, E., Fuentealba, L. C. & De Robertis, E. M. Integrating positional information at the level of Smad1/5/8. *Curr. Opin. Genet. Dev.* **18**, 304–310 (2008).

doi:10.1038/nature08306

ARTICLES

Structures of the tRNA export factor in the nuclear and cytosolic states

Atlanta G. Cook¹, Noemi Fukuhara^{2,†}, Martin Jinek^{2,†} & Elena Conti^{1,2}

Transfer RNAs are among the most ubiquitous molecules in cells, central to decoding information from messenger RNAs on translating ribosomes. In eukaryotic cells, tRNAs are actively transported from their site of synthesis in the nucleus to their site of function in the cytosol. This is mediated by a dedicated nucleo-cytoplasmic transport factor of the karyopherin- β family (Xpot, also known as Los1 in *Saccharomyces cerevisiae*). Here we report the 3.2 Å resolution structure of *Schizosaccharomyces pombe* Xpot in complex with tRNA and RanGTP, and the 3.1 Å structure of unbound Xpot, revealing both nuclear and cytosolic snapshots of this transport factor. Xpot undergoes a large conformational change on binding cargo, wrapping around the tRNA and, in particular, binding to the tRNA 5' and 3' ends. The binding mode explains how Xpot can recognize all mature tRNAs in the cell and yet distinguish them from those that have not been properly processed, thus coupling tRNA export to quality control.

During interphase, tRNAs relocate from the sites of transcription in the nucleus to the sites of translation in the cytosol^{1,2}. There they decode the genetic information in mRNAs by transferring cognate amino acids on translating ribosomes. tRNAs are exported out of the nucleus through the nuclear pore complexes (NPCs), large assemblies that puncture the nuclear envelope^{3–5}. Although tRNAs are smaller than 40 kDa—the cutoff size for passive diffusion through the NPC—a transport receptor is required for the tRNA to overcome the permeability barrier and hydrophobic meshwork of the NPC⁶. The main nucleo-cytoplasmic transport factor for tRNAs was identified ten years ago as the protein exportin-t (Xpot) in vertebrates and its orthologue Los1 in yeast^{7–9}.

Xpot is a member of the karyopherin- β family and is regulated by the small GTPase Ran^{10,11}. Ran has two functional states: the GTP-bound form (RanGTP) that is present at high concentrations in the nucleus, and the GDP-bound form that predominates in the cytoplasm. Xpot binds to tRNA in the presence of RanGTP in the nucleus. After NPC translocation, the ternary cargo-bound complex is disassembled at the cytoplasmic side of the nuclear envelope by the hydrolysis of Ran-bound GTP. In vertebrates, tRNA export can also be mediated by exportin-5, the export factor for pre-microRNAs¹². However, although exportin-5 can transport both pre-microRNAs and tRNAs, Xpot is exclusively dedicated to tRNAs¹³.

tRNAs fold into an L-shaped tertiary structure^{14,15}. One arm features the terminal acceptor stem that contains the amino-acid attachment site (acceptor arm), whereas the other arm contains the anticodon loop that base pairs to the corresponding codon in the mRNA (anticodon arm). The two arms are joined by the D, T Ψ C and variable loops that form the 'corner' or 'elbow' of the L-shaped molecule. This mature, functional form of tRNA ranges between 75 and 95 nucleotides in length and is the product of several processing reactions on the primary transcript. Pre-tRNA precursors are transcribed with extensions at the 5' and 3' ends that are cleaved by specific RNases^{16,17}. The processed 5' end terminates with a monophosphoryl group, whereas the 3' end terminates with a single base overhang, to which the moiety for amino-acid attachment (the terminal CCA motif) is subsequently

added¹⁸. Several families of pre-tRNAs contain an intervening sequence in the anticodon arm that is spliced out, usually before the maturation of the termini¹⁹. Finally, bases can be modified, for example, by methylation or by conversion of uridine to pseudouridine^{20,21}.

Xpot is able to recognize all mature cellular tRNAs, provided that they are properly folded. Mutations that disrupt the tRNA tertiary structure impede transport, suggesting that Xpot recognizes a structural motif in the nucleic acid cargo^{19,22,23}. RNA protection assays have mapped the binding region to the tRNA acceptor arm and the T Ψ C loop, but, in contrast to aminoacyl tRNA synthetases, an acceptor-T Ψ C minihelix is not sufficient for Xpot binding²². In addition, mature 3' and 5' ends are a prerequisite for efficient tRNA export, and splicing or base modifications seem not to be essential^{18,19,22}. In this work, we address the structural basis by which the nuclear export factor Xpot is able to distinguish between unprocessed pre-tRNAs and mature tRNA.

Structure determination

We determined the crystal structures of Xpot in its nuclear export complex with RanGTP and tRNA and in the unbound form corresponding to the cytosolic state. We obtained diffracting crystals using Xpot from fission yeast, which shares about 24% sequence identity with the human and *S. cerevisiae* orthologues (Supplementary Fig. 1). The structure of unbound Xpot was solved by multiple isomorphous replacement with anomalous signal (MIRAS) using crystals derivatized with gold, mercury and selenomethionine. The final model has been refined against a native data set at 3.1 Å resolution to an R_{free} of 31.3% and R_{work} of 27.3%, and includes 911 amino-acid residues out of 978. A summary of data and refinement statistics can be found in Supplementary Table 1. Although the electron density of unbound Xpot was well-defined for the N-terminal two-thirds of the molecule, most of the side chains in the carboxy-terminal third could not be modelled. The sequence was assigned using the positions of the selenium atoms and was later verified with the model of Xpot in the tRNA–RanGTP complex, in which the C terminus of the molecule is better ordered.

¹Structural Cell Biology, MPI for Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany. ²EMBL, Meyerhofstrasse 1, Heidelberg D69117, Germany. [†]Present addresses: Division of Cell and Molecular Biology, Biophysics section, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK (N.F.); Department of Molecular Cell Biology, University of California at Berkeley, 731 Stanley Hall, Berkeley, California 94720, USA (M.J.).

S. pombe Xpot was crystallized in complex with the minimal binding portions of RanGTP and Phe-tRNA from budding yeast. Ran was truncated to contain residues 8–179, excluding the C-terminal region (residues 180–219) that is disordered or absent in the crystal structures of other karyopherin- β -RanGTP complexes^{24–27}. In addition, the Gln71Leu mutation was introduced to limit GTP hydrolysis²⁸. Yeast Phe-tRNA was engineered on the basis of biochemical experiments that showed that the tRNA anticodon arm is accessible to RNases in the ternary export complex and is therefore likely to be exposed²². A truncated version of Phe-tRNA was transcribed in which most of the anticodon arm had been deleted and replaced with a UUCG tetraloop²⁹ (Supplementary Fig. 2). Crystals of this ternary complex diffracted to 3.2 Å resolution and contained two complexes per asymmetric unit. The structure was solved by molecular replacement using the coordinates of Ran (Protein Data Bank (PDB) accession 1IBR (ref. 30), tRNA PDB accession 1EVV (ref. 14)) and fragments of unbound Xpot. The structure has been refined to an R_{free} of 29.4% and R_{work} of 24.1%. The final model contains 941 residues of Xpot and is missing a few disordered loops. The two copies of Xpot in the asymmetric unit have a root mean square deviation (r.m.s.d.) of 0.67 Å over the C α atoms from residues 2 to 896. The last four helices of the structure are less well-defined and show a relative shift between the two copies in the asymmetric unit (r.m.s.d. over all Xpot C α atoms is 1.02 Å). Data collection and refinement statistics can be found in Supplementary Table 2. The quality of the electron density is shown in Supplementary Fig. 3 for regions discussed in the text and the B-factor distributions are shown in Supplementary Fig. 4.

RanGTP interactions with Xpot

The overall structure of Xpot forms a large superhelix that contains 19 tandem HEAT repeats (Fig. 1 and Supplementary Fig. 5). HEAT repeats are ~40 amino-acid motifs that typically fold into two α -helices, A and B. The repeats pack against each other, generally with a ~15° clockwise rotation between successive repeats³¹. This leads to a relatively regular superhelical structure with the A helices forming the outer convex surface and the B helices forming the inner concave surface. The structure can be considered to be composed of two ‘arches’, an amino-terminal arch comprising HEAT repeats 1–9 (residues 1–427) and a C-terminal arch defined from HEAT repeat 10 to the C terminus. The regularity of the Xpot superhelix is interrupted by a kink between HEATs 9 and 10 and another between HEATs 13 and 14, where the rotation is anticlockwise. The last three helices of the structure do not form a HEAT motif but instead zigzag across the face of HEAT 19, capping the superhelix (Supplementary Fig. 5). Xpot contains several long insertions, either in loops between successive HEAT repeats (inter-repeat) or between the A and B helices of the same repeat (intra-repeat). In particular, a loop is

inserted between HEATs 4 and 5 that contains an extra helix, a long loop is present within HEAT 9, and a third insert is placed between HEATs 10 and 11 (Supplementary Fig. 5). Finally, HEAT 13 contains a large intra-repeat insertion whereby the A helix ends in an extended strand that packs against the elongated B helix.

Both RanGTP and tRNA engage in extensive interactions with Xpot as well as with each other, burying a total surface of 2,640 Å² on Xpot and an average of 355 Å² between Ran and tRNA (Fig. 1). RanGTP interacts primarily with the N-terminal arch of Xpot, similar to what has been observed with other karyopherins^{24–27}. In particular, HEATs 1 to 4 interact with the switch II region of Ran (residues 68 to 81) (Fig. 2 and Supplementary Fig. 6). A second interaction involves the large intra-repeat loop of HEAT 9, which contains several aspartate and glutamate residues that interact with Lys 161, Tyr 149 and Asn 158 of Ran (Fig. 2). This insertion has a similar role in binding RanGTP as the ‘acidic loop’ found in the structures of importin β and transportin, although in these cases the acidic insertion is found at HEAT 8 (refs 24, 25, 30 and 32). The C-terminal arch of Xpot interacts with Ran at the extended intra-repeat loop of HEAT 13 and at the intra-repeat loop of HEAT 17, which contacts the switch I region (Fig. 2). Other exportins also contact the switch I region of Ran by the C-terminal arch, in the case of Cse1 by an extended intra-repeat loop of HEAT 19 (ref. 26) and in the case of Crm1 by an extensive interaction interface formed by the intra-repeat loops of HEATs 17 to 19 (ref. 27). Although the interactions of RanGTP with the C-terminal arch of Xpot are different to those observed with other exportins, the overall recognition mechanism is similar in that Xpot interacts with both the switch I and switch II regions of RanGTP and directly probes the nucleotide-bound state of Ran.

Recognition of tRNA by Xpot

Xpot wraps around tRNA with contacts distributed around the acceptor arm, T Ψ C loop and D loop (Figs 1 and 3a). The truncated anticodon arm extends into solvent and makes no contacts with the protein. Indeed, we observed only partial electron density for the tetraloop at the tip of the tRNA anticodon arm, which has therefore not been modelled. The acceptor arm binds at an extended, positively charged surface of Xpot formed by the B helices from HEATs 8 to 18 (Supplementary Fig. 5 and Fig. 3b). Most of the positively charged residues lining this surface are evolutionarily conserved (Fig. 3c), in particular at the B helix of HEAT 12 (Supplementary Fig. 1). Indeed, mutations of the equivalent HEAT 12B residues in human XPOT were reported to cause a loss of tRNA binding *in vitro*³³. tRNA also contacts Ran. The basic surface patch of Ran (consisting of residues Lys 129, Arg 131, Lys 134, Lys 136, Lys 97, Lys 101 and Lys 143) interacts with the backbone phosphates of tRNA at nucleotides 2–4 and 63–65 (for clarity, standard numbering of yeast Phe-tRNA has been used³⁴; Supplementary Fig. 2).

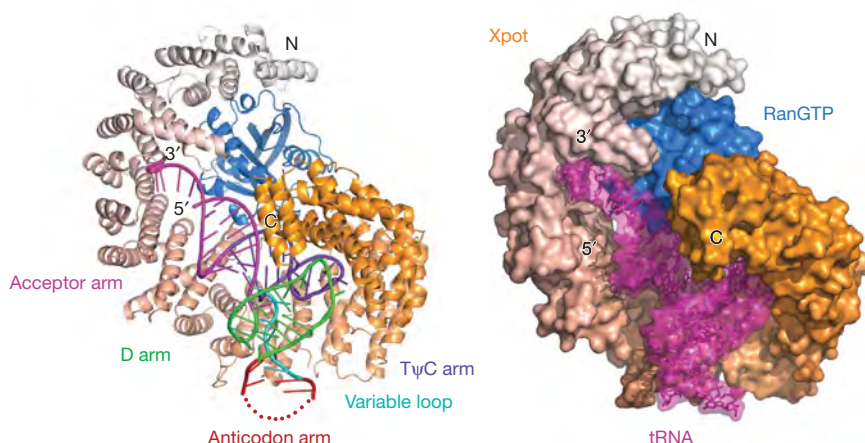


Figure 1 | Structure of the Xpot-RanGTP-tRNA complex. View of the complex as ribbon (left) and surface (right) representations. Xpot is shown with a colour gradient from grey (N terminus) to orange (C terminus). RanGTP is in blue, tRNA is shown in magenta on the right, and on the left the different regions are colour-coded and indicated accordingly. The anticodon arm projects into solvent and is only partially visible. Pictures were generated with PyMOL (Delano Scientific).

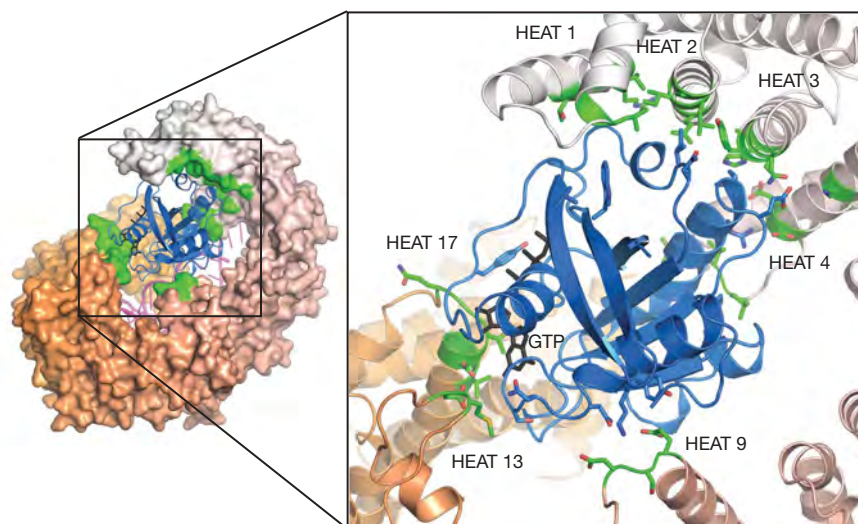


Figure 2 | The binding of RanGTP to Xpot. The ternary export complex is viewed with a 180° clockwise rotation around the vertical axis relative to the view in Fig. 1. On the left, Xpot is shown as a surface representation coloured from grey (N terminus) to orange (C terminus), whereas Ran (at the front) and tRNA (at the back) are shown in a ribbon representation in blue and magenta. The portions of Xpot that interact with Ran are highlighted in green. On the right, a magnified view shows the interactions between the Xpot HEAT repeats and Ran. GTP is shown in ball-and-stick representation (black).

In the complex, tRNA positions the 5' and 3' ends of the acceptor stem across the N-terminal arch of Xpot, whereas the 'corner' of the L-shaped structure is placed in the tightly curved surface of the Xpot C-terminal arch (HEATs 10 to 19) (Fig. 3a). Here, His 766 and Gln 763 of Xpot pack against the bases of G19 and C56, which link the D and TΨC loops. Disruption of this base pair prevents nuclear export, probably by compromising the proper folding of the tRNA 'corner'^{22,23}. The 5' end of tRNA is buried between a part of the N-terminal arch of Xpot (HEATs 6 to 8) and the basic patch of Ran (Fig. 4a). Specificity in the recognition of the 5' end seems not to rely on direct contacts with either Xpot or Ran, but rather on steric

constraints, as the geometry of the binding pocket for the 5' end is incompatible with the presence of a 5' extension.

The CCA sequence at the 3' end of tRNA (nucleotides 74–76) lies in a groove created by the inter-repeat sequences between HEATs 4 and 7 (Fig. 4b). Three highly conserved residues of Xpot (Lys 177, Arg 181 and Lys 259) line the groove and interact with the phosphate groups of the 3' overhang. The two cytosines of the CCA motif engage in direct interactions with Lys 260. The side chain of this lysine contacts the C74 ring (at the nitrogen in position 3) and the main-chain carbonyl oxygen contacts the ribose of C75 (at the 2' hydroxyl). The last base, A76, twists away from Xpot and stacks

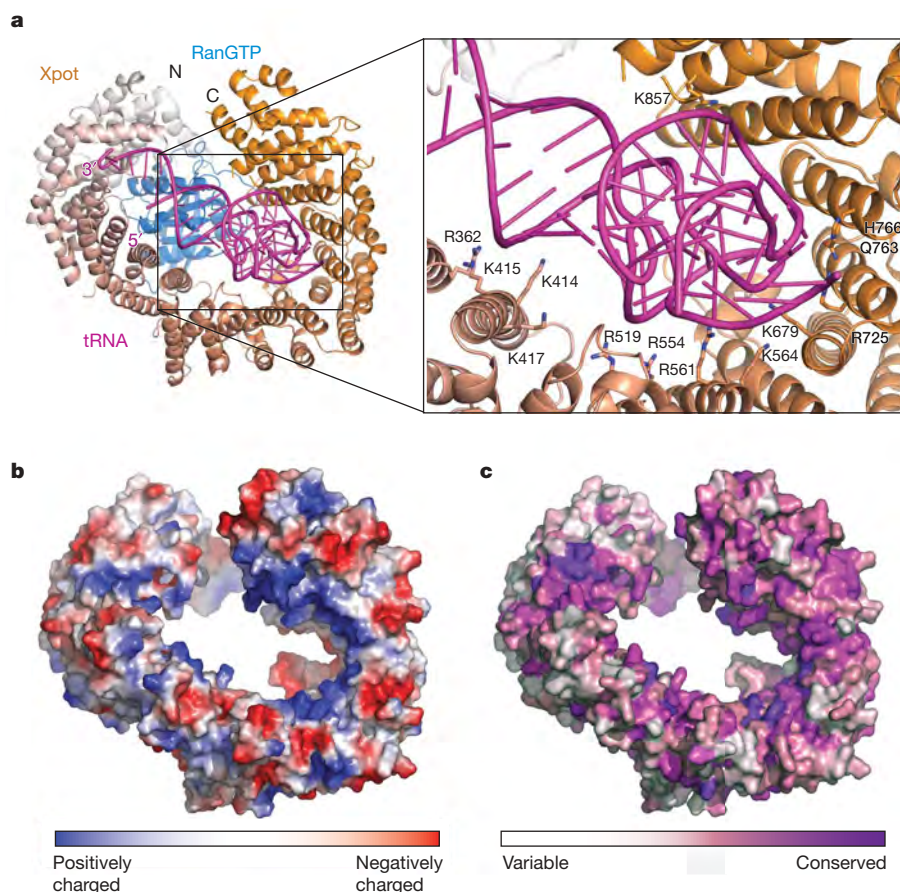


Figure 3 | Complementary tRNA-binding surfaces on Xpot. **a**, The export complex is shown after a 50° anticlockwise rotation around the horizontal axis relative to the view in Fig. 1. tRNA (magenta) is at the front. The magnified view (right) shows the residues of Xpot that contact tRNA. **b**, **c**, Xpot is rendered as a surface in the same orientation as in **a**, with RanGTP and tRNA removed for clarity. The surfaces are coloured according to electrostatic potential (with positively charged residues in blue, **b**) and according to conservation (with a gradient from white to violet indicating increasingly conserved residues, **c**).

against C75. The adenosine is solvent exposed, with the 3' OH placed close to the highly conserved Asp 178. The proximity of Asp 178 to the 3' OH would probably repel the phosphate group of an extra nucleotide but would be compatible with the presence of an aminoacyl group of a charged tRNA¹⁹ (Supplementary Fig. 7). Finally, the helical insertion between HEATs 4 and 5 contacts not only the RNA backbone of the 3' end but also Ran, and might therefore provide a point of communication between tRNA and RanGTP binding.

Cargo-induced conformational changes

The structure of RanGTP in the Xpot ternary complex is very similar to that reported in other karyopherin- β complexes, with an r.m.s.d. of about 0.5 Å for all C α atoms^{24–27}. Comparison of tRNA with the high-resolution structure of Phe-tRNA¹⁴ shows an r.m.s.d. of 1.9 Å for all atoms encompassing nucleotides 1 to 73, reflecting a slight rotation of the acceptor arm with respect to the anticodon arm. In contrast to these minor differences, Xpot shows a marked conformational change when the cargo-bound and unbound structures are compared (Fig. 5a). In the structure of Xpot in the unbound state, the karyopherin superhelix is in an extended conformation, measuring about 111 Å along the axis of the superhelix (calculated as the distance between residues 52 and 922). Xpot is rather more compact in the ternary complex (81 Å along the superhelical axis): the C terminus is displaced by 30 Å closer to the N terminus as compared to the unbound structure (Fig. 5a). The observed differences in the two crystal structures agree well with results from solution studies by small angle X-ray scattering³⁵.

The difference in conformation observed in Xpot is not the result of discrete hinge movements as was the case in the karyopherin Cse1 (refs 26, 36), but is the cumulative effect of small changes distributed

over the entire molecule. The largest differences are at the regions around the Ran-binding loops of HEAT repeats 9, 13 and 17 (Supplementary Fig. 8). A partial melting of the helix A in HEAT 7 introduces a kink that moves helix 7B towards the concave face of the superhelix. This rearrangement in turns displaces HEAT repeats 8 and 9 in the same direction, creating a tighter turn of the superhelix and allowing the acidic loop on HEAT 9 to approach RanGTP. A similar displacement towards the concave surface is observed at HEAT 16, again creating a tighter superhelical turn and allowing the loop region of HEAT 17 to interact with RanGTP (Supplementary Fig. 8). Thus, although the most extensive Ran-binding surface is at the N-terminal arch, the Ran-binding loops in the central and C-terminal portion of Xpot seem to have a crucial role in forming the tight curvature of the molecule. This observation from the structure predicts that the binding of Ran to Xpot might be sufficient to prime the surface of the karyopherin into the curved conformation that is complementary to the shape of tRNA. This prediction can be recapitulated biochemically: RanGTP can form a stable binary complex with Xpot (Fig. 5b, lane 3), whereas tRNA can only bind to Xpot when RanGTP is present (Fig. 5b, lane 4, compare with lane 2). The observation that *S. pombe* Xpot can form a stable complex with RanGTP in the absence of tRNA suggests that the ternary, cargo-bound complex may proceed through an Xpot–RanGTP intermediate.

The mechanism of assembly and disassembly of cargo complexes by *S. pombe* Xpot seems to be qualitatively different from that observed in other export complexes such as Cse1 and Crm1 (refs 26, 27, 36 and 37; Supplementary Figs 9 and 10). In Cse1, a large conformational change is also observed between the bound and unbound states, but in the opposite direction as compared to Xpot: Cse1 is more compact in the

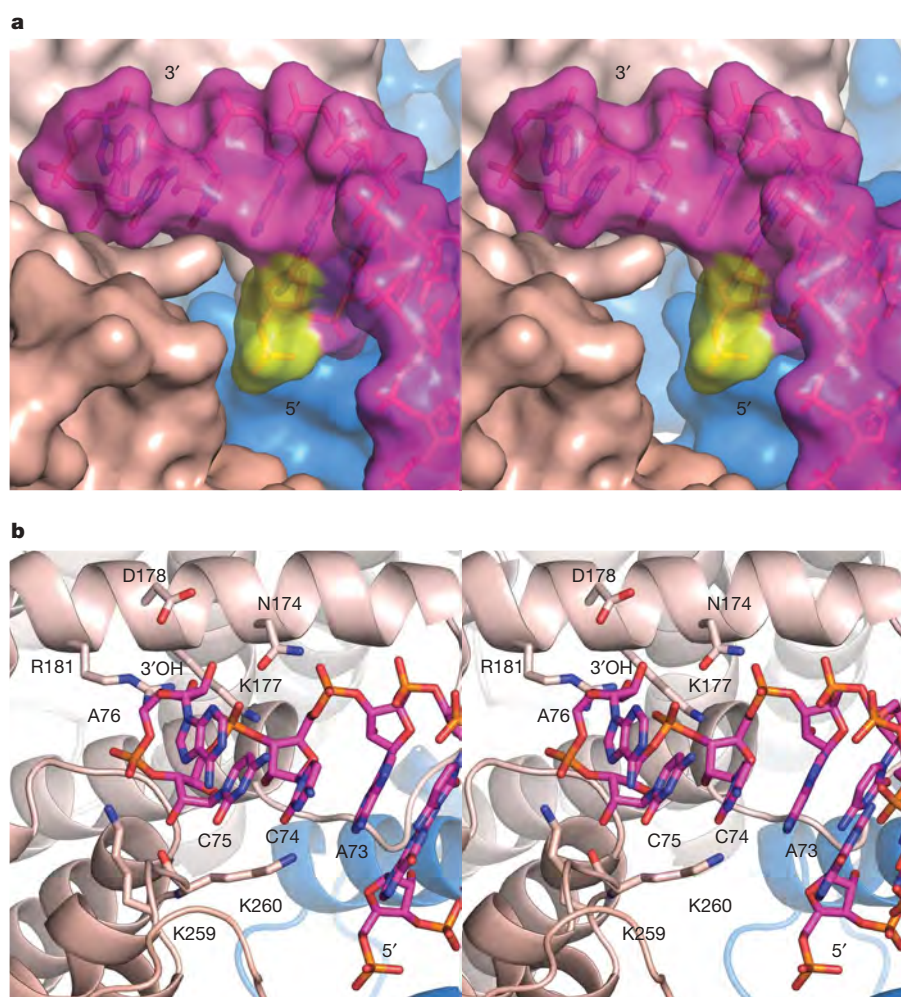


Figure 4 | Recognition of the tRNA 3' and 5' ends. **a**, Stereo view of the surface representation of the tRNA 3' and 5' ends in a similar orientation and colour-coding as in Fig. 3. The monophosphoryl group of the tRNA 5' end (yellow) is buried in the export complex. **b**, Stereo diagram showing the interactions between Xpot and the CCA motif at the tRNA 3' end (C74, C75 and A76). The tRNA is in a ball-and-stick representation, with carbon atoms in magenta. Xpot residues discussed in the text are highlighted. The figure also shows the tRNA monophosphoryl group at the 5' end.

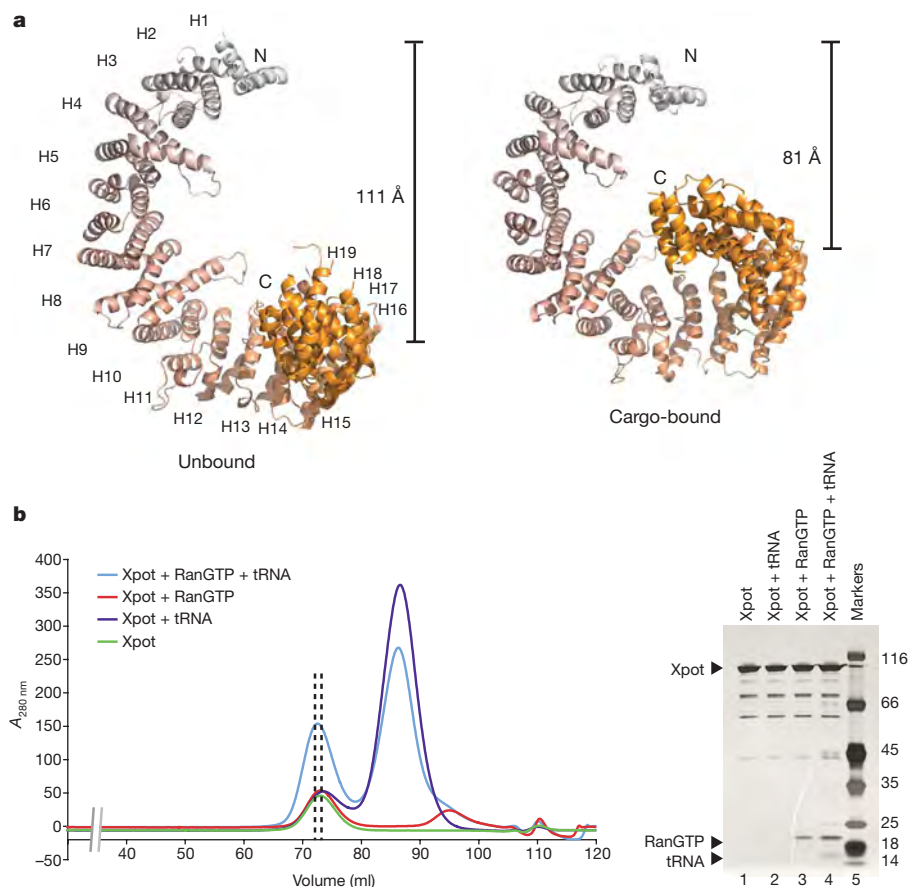


Figure 5 | Conformational change of Xpot unbound and in complex. **a**, Xpot shown in the unbound state (left) versus the cargo-bound state (right, with RanGTP and tRNA removed for clarity). The molecules have the same colour-coding and view as in Fig. 1, after superposition of the Xpot N-terminal arches. **b**, Overlay of the size-exclusion chromatograms of Xpot alone or with ligands (in excess). Samples from the peak fraction (indicated by dotted lines) are shown on an SDS-PAGE silver-stained gel (right). tRNA stains less intensely than proteins. Minor bands between 45 and 100 kDa are degradation impurities of Xpot. Molecular mass markers are shown in lane 5.

unbound state than in the ternary export complex^{26,35,36}. Cargo binding in the cytosol by Cse1 is prevented by the presence of an intramolecular interaction that constrains the karyopherin superhelix in a tight ring-like structure, distorting the cargo-binding surface. In Xpot, the open conformation of the unbound state also distorts the cargo-binding surface, in this case by an outward displacement rather than an inward compaction of the individual contact points for cargo recognition.

Conclusions

Xpot is a general transporter of tRNAs and as such recognizes structural features that are common to all tRNAs. The structure of the complex of Xpot with RanGTP and Phe-tRNA presented here shows how this is achieved. The C-terminal arch of Xpot has a tight curvature that complements the overall shape and electrostatic character of the tRNA and binds to the composite surface created by the acceptor arm, TΨC loop and D loop. The binding surface of Xpot is lined with basic residues that interact with the phospho-ribose backbone of the tRNA without making base-specific interactions. This mode of recognition is different from that seen in aminoacyl tRNA synthetases where a substantial surface on both acceptor and anticodon arms is usually recognized, but generally only on one face of the tRNA³⁸.

Xpot also makes specific interactions with discrete features of tRNAs, in particular with the 3' single-stranded overhang of the CCA motif. Although there is no direct read-out of the sequence of the CCA motif (apart from base C74), the binding site on Xpot constrains this sequence to be single stranded and no more than four bases long. A conserved aspartate is found in close proximity to the 3' hydroxyl moiety of tRNA and could have a role in recognizing amino-acyl groups of charged tRNAs. The 5' end of tRNA is not contacted directly by either RanGTP or Xpot. However, this group is buried in a cleft between the transport factor and the GTPase so that extensions to the 5' end would clash with Xpot. These steric constraints on the binding of the 5' and 3' ends of the tRNA underpin the mechanism by which Xpot discriminates against pre-tRNAs with

extended 5' and 3' termini. tRNAs with incomplete 3' ends, that is, lacking the CCA motif, would be too short to interact with the N-terminal arch of Xpot. The structural framework of Xpot therefore provides a quality control step, ensuring that only mature tRNAs are exported.

Xpot is found in both the unbound cytosolic form, with an open three-dimensional structure, and in the cargo-bound nuclear form that is more compact. The tRNA-interacting residues are exposed and accessible in unbound Xpot. Yet, the lack of overall shape complementarity in the absence of RanGTP probably plays an important part in the disassembly of the ternary complex in the cytosol and in preventing a potential re-import of tRNAs. Along these lines, it is possible that the assembly process of the complex in the nucleus might occur in a stepwise fashion with the formation of a RanGTP–Xpot binary intermediate. The emerging picture from the structural characterization of different karyopherins is that the same basic architecture is used in different ways to solve the problem of cargo binding and release. On a more general level, the Xpot structure shows how the HEAT repeat architecture, that is usually associated with protein–protein interactions, is recruited to recognize a nucleic acid.

METHODS SUMMARY

S. pombe Xpot was expressed as a glutathione *S*-transferase (GST)-fusion protein in *Escherichia coli* BL21-CodonPlus GOLD (DE3) RIL cells (Stratagene) using an overnight induction at 18 °C. The protein was purified using glutathione sepharose, followed by GST cleavage with 3C protease, heparin sepharose and size-exclusion chromatography. Xpot was concentrated to 8 mg ml⁻¹ and crystallized by vapour diffusion at 18 °C using a reservoir solution containing 100 mM MES, pH 6.5, 30% (v/v) PEG400 and 160 mM CaCl₂.

The *S. cerevisiae* Ran (residues 8–179) Gln71Leu mutant was expressed as a TEV-cleavable His-tagged protein in *E. coli* DL-41 cells³⁹. The protein was purified using a cobalt affinity resin and the tag was removed with TEV protease. Ran was loaded with GTP and purified by cation-exchange chromatography on a ResourceS column (GE Healthcare) to remove nucleic-acid impurities. tRNA

was cloned into pUC19 and produced by 'run-off' *in vitro* transcription. The sample was desalted and tRNA was separated on a Superdex 75 gel filtration column in the presence of 4 mM MgCl₂. For complex formation, Xpot was incubated with a threefold molar excess of RanGTP and tRNA. The ternary complex was purified to homogeneity by size-exclusion chromatography in 20 mM sodium acetate, pH 5.0, 100 mM NaCl and 4 mM MgCl₂. It was concentrated to 15 mg ml⁻¹ and crystallized by vapour diffusion at 18 °C using 100 mM Bis-Tris, pH 5.5, 20% (w/v) PEG3350, 180 mM sodium acetate in the reservoir solution.

Diffraction data were collected at the Swiss Light Source (Villigen, Switzerland) and the structures were determined by MIRAS and molecular replacement. Data and refinement statistics are summarized in Supplementary Tables 1 and 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 June; accepted 6 August 2009.

Published online 13 August 2009.

- Kohler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nature Rev. Mol. Cell Biol.* **8**, 761–773 (2007).
- Rodriguez, M. S., Dargemont, C. & Stutz, F. Nuclear export of RNA. *Biol. Cell* **96**, 639–655 (2004).
- Elad, N., Maimon, T., Frenkiel-Krispin, D., Lim, R. Y. & Medalia, O. Structural analysis of the nuclear pore complex by integrated approaches. *Curr. Opin. Struct. Biol.* **19**, 226–232 (2009).
- Lim, R. Y., Aebi, U. & Fahrenkrog, B. Towards reconciling structure and function in the nuclear pore complex. *Histochem. Cell Biol.* **129**, 105–116 (2008).
- Peters, R. Translocation through the nuclear pore: Kaps pave the way. *Bioessays* **31**, 466–477 (2009).
- Zaslloff, M. tRNA transport from the nucleus in a eukaryotic cell: carrier-mediated translocation process. *Proc. Natl Acad. Sci. USA* **80**, 6436–6440 (1983).
- Arts, G. J., Fornerod, M. & Mattaj, I. W. Identification of a nuclear export receptor for tRNA. *Curr. Biol.* **8**, 305–314 (1998).
- Kutay, U. *et al.* Identification of a tRNA-specific nuclear export receptor. *Mol. Cell* **1**, 359–369 (1998).
- Hellmuth, K. *et al.* Yeast Los1p has properties of an exportin-like nucleocytoplasmic transport factor for tRNA. *Mol. Cell Biol.* **18**, 6374–6386 (1998).
- Cook, A., Bono, F., Jinek, M. & Conti, E. Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* **76**, 647–671 (2007).
- Madrid, A. S. & Weis, K. Nuclear transport is becoming crystal clear. *Chromosoma* **115**, 98–109 (2006).
- Calado, A., Treichel, N., Muller, E. C., Otto, A. & Kutay, U. Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *EMBO J.* **21**, 6216–6224 (2002).
- Shibata, S. *et al.* Exportin-5 orthologues are functionally divergent among species. *Nucleic Acids Res.* **34**, 4711–4721 (2006).
- Jovine, L., Djordjevic, S. & Rhodes, D. The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg²⁺ in 15-year old crystals. *J. Mol. Biol.* **301**, 401–414 (2000).
- Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
- Marvin, M. C. & Engelke, D. R. RNase P: increased versatility through protein complexity? *RNA Biol.* **6**, 40–42 (2009).
- Spath, B., Canino, G. & Marchfelder, A. tRNase Z: the end is not in sight. *Cell. Mol. Life Sci.* **64**, 2404–2412 (2007).
- Xiong, Y. & Steitz, T. A. A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. *Curr. Opin. Struct. Biol.* **16**, 12–17 (2006).
- Lund, E. & Dahlberg, J. E. Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science* **282**, 2082–2085 (1998).
- Bjork, G. R. *et al.* Transfer RNA modification. *Annu. Rev. Biochem.* **56**, 263–287 (1987).
- Iwata-Reuyl, D. An embarrassment of riches: the enzymology of RNA modification. *Curr. Opin. Chem. Biol.* **12**, 126–133 (2008).
- Arts, G. J., Kuersten, S., Romby, P., Ehresmann, B. & Mattaj, I. W. The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J.* **17**, 7430–7441 (1998).
- Lipowsky, G. *et al.* Coordination of tRNA nuclear export with processing of tRNA. *RNA* **5**, 539–549 (1999).
- Chook, Y. M. & Blobel, G. Structure of the nuclear transport complex karyopherin-β2-Ran-GppNHp. *Nature* **399**, 230–237 (1999).
- Lee, S. J., Matsuura, Y., Liu, S. M. & Stewart, M. Structural basis for nuclear import complex dissociation by RanGTP. *Nature* **435**, 693–696 (2005).
- Matsuura, Y. & Stewart, M. Structural basis for the assembly of a nuclear export complex. *Nature* **432**, 872–877 (2004).
- Monecke, T. *et al.* Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* **324**, 1087–1091 (2009).
- Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A. & Ponstingl, H. RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc. Natl Acad. Sci. USA* **91**, 2587–2591 (1994).
- Bloomfield, V. A., Crothers, D. N. & Tinoco, I. *Nucleic Acids: Structures, Properties and Functions* Ch. 8 298–300 (Macmillan, 1999).
- Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D. & Wittinghofer, A. Structural view of the Ran-Importin β interaction at 2.3 Å resolution. *Cell* **97**, 635–646 (1999).
- Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W. & Bork, P. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1–18 (2001).
- Cingolani, G., Petosa, C., Weis, K. & Müller, C. W. Structure of importin-β bound to the IBB domain of importin-α. *Nature* **399**, 221–229 (1999).
- Kuersten, S., Arts, G. J., Walther, T. C., Englmeier, L. & Mattaj, I. W. Steady-state nuclear localization of exportin-t involves RanGTP binding and two distinct nuclear pore complex interaction domains. *Mol. Cell Biol.* **22**, 5708–5720 (2002).
- Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148–153 (1998).
- Fukuhara, N., Fernandez, E., Ebert, J., Conti, E. & Svergun, D. Conformational variability of nucleocytoplasmic transport factors. *J. Biol. Chem.* **279**, 2176–2181 (2004).
- Cook, A. *et al.* The structure of the nuclear export receptor Cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding. *Mol. Cell* **18**, 355–367 (2005).
- Dong, X., Biswas, A. & Chook, Y. M. Structural basis for assembly and disassembly of the CRM1 nuclear export complex. *Nature Struct. Mol. Biol.* **16**, 558–560 (2009).
- Moras, D. Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.* **17**, 159–164 (1992).
- LeMaster, D. M. & Richards, F. M. 1H–15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–7268 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank P. Brick for comments and critical reading of the manuscript. We also thank P. Reichelt and J. Ebert for technical assistance, S. Kuersten and I. Mattaj at the initial phase of the project, K. Valer-Saldana, S. Pleyer and J. Basquin of the MPI-Martinsried crystallization facility and the staff at the Swiss Light Source (Villigen, Switzerland) for assistance during data collection. We would like to thank C. Vornrhein (Global Phasing) for optimising the phase calculations of unbound Xpot with SHARP. This study was supported by the Max Planck Gesellschaft, the European Molecular Biology Laboratory (EMBL), the Sonderforschungsbereich SFB646 and the Gottfried Wilhelm Leibniz Program of the Deutsche Forschungsgemeinschaft (DFG), the EU grant 3D Repertoire contract number LSHG-CT-2005-512028.

Author Contributions N.F. identified *S. pombe* Xpot as the most promising orthologue for crystallisation, created the truncated tRNA construct and obtained initial, low-resolution crystals for the cargo-bound complex. M.J. crystallised unbound Xpot. A.G.C. solved the structure of unbound Xpot, obtained well-diffracting crystals of the ternary complex and solved its structure. E.C. supervised the project. A.G.C. and E.C. wrote the paper.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession codes 3IBV for Xpot and 3ICQ for the Xpot–RanGTP–tRNA complex. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.C. (conti@biochem.mpg.de).

METHODS

Protein expression and purification. *S. pombe* exportin-t (Xpot) was PCR amplified from genomic DNA. The full-length insert was cloned into a modified pGEX vector and expressed as a GST-fusion protein cleavable with the rhinovirus 3C protease⁴⁰. *S. pombe* and *S. cerevisiae* Ran share 87% sequence identity overall and 91% sequence identity over residues 8–179 of yeast Ran, the construct used in this study. Residues 8–179 of *S. cerevisiae* Ran (Gsp1) were subcloned in a pProEX-HTb expression vector (Life Sciences), with an N-terminal 6×-His tag and a tobacco etch virus (TEV) protease site. The Gln71Leu mutation was introduced into this construct using QuikChange mutagenesis according to the manufacturer's instructions (Stratagene).

Xpot was expressed in *E. coli* BL21-CodonPlus GOLD (DE3) RIL cells (Stratagene) using an overnight induction at 18 °C. The protein was purified by batch binding to glutathione sepharose resin (GE Healthcare) in a buffer containing 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mM dithiothreitol (DTT) and eluted by cleavage of the GST tag with 3C protease. The protein was further purified with heparin sepharose and a final size-exclusion chromatography step on a Superdex 200 HR 16/60 column (GE Healthcare) in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 1 mM DTT. For phasing, selenomethionine-labelled Xpot was expressed in B834 cells grown in minimal media supplemented with amino acids and selenomethionine⁴¹. Selenomethionine-derivatized protein was purified using the same protocol as for the unmodified protein.

Ran was expressed in *E. coli* DL-41 cells³⁹ with an overnight induction at 18 °C. The protein was purified using a metal affinity resin charged with Co²⁺ ions (Clontech). The hexahistidine tag was removed with TEV protease and a second passage over the Co²⁺ resin (using 20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 4 mM MgCl₂, 10% glycerol and 0.5 mM β-mercaptoethanol) removed any remaining uncut protein and TEV protease. Ran was then incubated with 7 mM EDTA to remove divalent metal ions and to strip out the bound nucleotide, and dialysed (in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 4 mM MgCl₂, 10% glycerol and 1 mM DTT). It was then incubated with a fivefold excess of GTP and 10 mM MgCl₂ at room temperature for 30 min to produce GTP-loaded Ran. As a final purification step, RanGTP was purified by cation-exchange chromatography (ResourceS, GE Healthcare).

RNA preparation. Truncated yeast Phe-tRNA was transcribed *in vitro* from a T7 promoter. Overlapping DNA oligonucleotides encoding the T7 promoter and the truncated sequence were annealed and extended using the DNA Polymerase I Klenow fragment (New England Biolabs). The extended sequence was then cloned into pUC19 (New England Biolabs) and amplified to produce templates for *in vitro* transcription. The purified plasmid was linearized with MvaI (Fermentas). The restriction enzyme was removed using phenol extraction followed by ethanol precipitation. The truncated tRNA was produced by run-off transcription in a buffer containing 40 mM Tris-HCl, pH 8.0, 5 mM DTT, 28 mM MgCl₂, 1 mM spermidine, 0.01% Triton X-100, 0.1 mg ml⁻¹ linearized template, 4 mM of each NTP, 80 mM GMP and 0.5 mg ml⁻¹ T7 polymerase. GMP was added in a large excess to ensure a monophosphate group at the 5' end of tRNA. The reaction was allowed to proceed at 37 °C for 2 h, after which the salt precipitate was removed by centrifugation and the reaction was allowed to proceed for a further 2 h. The DNA template was digested for 30 min using RNase-free DNase (New England Biolabs). Excess nucleotides were removed on a desalting column (GE Healthcare) in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). tRNA was separated on a Superdex 75 gel filtration column (GE Healthcare) and eluted in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 4 mM MgCl₂.

Crystallization of unbound Xpot and of Xpot–RanGTP–tRNA. Xpot was concentrated to 8 mg ml⁻¹ and crystallized using the hanging drop vapour diffusion method in 24-well VDX plates (Hampton Research). One-microlitre drops were set up on siliconized glass coverslips suspended over 300 µl of reservoir solution containing 100 mM MES, pH 6.5, 30% (v/v) PEG400 and 160 mM CaCl₂. Rectangular plate-like crystals appeared after a few days at 18 °C.

For complex formation, purified Xpot was mixed with a threefold molar excess of RanGTP and incubated on ice for approximately 10 min. MgCl₂ was added to tRNA to a final concentration of 20 mM before mixing a threefold molar excess of the RNA with the Xpot–RanGTP sample. The ternary complex was then separated from the excess RanGTP and tRNA by size-exclusion chromatography with a Superdex 200 HR 16/60 column in 20 mM sodium acetate, pH 5.0, 100 mM NaCl and 4 mM MgCl₂. The purified complex was concentrated to ~15 mg ml⁻¹ and crystallized using the sitting drop vapour diffusion method. Crystals grew after 1 week at 18 °C using as reservoir solution 100 mM Bis-Tris, pH 5.5, 20% (w/v) PEG3350 and 180 mM sodium acetate.

Structure determination and analysis. All diffraction data were collected at beamlines X06SA and X10SA at the Swiss Light Source (Villigen, Switzerland). The crystals of both the unbound and cargo-bound Xpot grew in a cryoprotected condition and were cooled directly in a nitrogen stream at 100 K. All data were

indexed and scaled using MOSFLM⁴² and SCALA⁴³ with the exception of the native data set for the unbound Xpot. The native data for unbound Xpot were collected using the PILATUS pixel detector at beamline X06SA (ref. 44) and were indexed and scaled using the XDS package⁴⁵. The crystals of unbound Xpot grew in an orthorhombic space group (*P*₂₁₂₁₂) with two molecules in the asymmetric unit. In this crystal form the non-crystallographic two-fold axis is parallel to the crystallographic *b* axis, resulting in non-crystallographic translational symmetry. Further data sets for unbound Xpot were collected from crystals soaked with AuCN (collected at the peak wavelength, 1.037 Å) or EMP (collected at 1.007 Å) or from crystals grown from selenomethionine-derivatized protein (collected at 0.979 Å). The heavy-atom substructure for the gold derivative was determined using SHELX⁴⁶. Phase calculations and refinement using the extra data sets were carried out using SHARP⁴⁷.

Cargo-bound Xpot also crystallized in an orthorhombic space group (*P*₂₁₂₁₂) with two molecules in the asymmetric unit. The *a* and *b* axes in this crystal form were almost equal in length. As a consequence, many of the crystals had some degree of pseudomerohedral twinning (between 4% and 25%). The native data set used for structure determination was collected from a non-pseudomerohedrally twinned crystal at a wavelength of 1.07 Å. The structure was solved by molecular replacement using PHASER⁴⁸. Fragments of the Xpot from residues 1 to 386 and residues 448 to 793 were used to locate the Xpot chain. The Ran search model was derived from the structure of canine Ran in complex with the N-terminal fragment of importin-β (PDB accession 1IBR) and trimmed to fit the yeast sequence using CHAINSAW⁴⁹. A partial model of the acceptor stem and TΨC loop of the yeast Phe-tRNA (PDB accession 1EVV) was used as a search model for the tRNA. The molecular graphics program O⁵⁰ was used for model building and CNS 1.2 (ref. 51) was used for refinement. For the unbound structure, strict NCS constraints were applied to all but the last round of refinement. For the cargo-bound Xpot, NCS restraints were used in the early stages of refinement but were removed in later rounds of refinement. The final models were assessed for quality by MOLPROBITY⁵². The electrostatic potential was calculated using PYMOL (Delano Scientific) and surface conservation was calculated using CONSURF⁵³. Data and refinement statistics are summarized in Supplementary Tables 1 and 2. Both models show good stereochemistry. The Ramachandran values for the unbound Xpot structure are: Most favoured 94.3%, Additionally allowed 5.5% and Disallowed 0.2%. The Ramachandran values for the Xpot–RanGTP–tRNA complex are: Most favoured 93.4%, Additionally allowed 6.1% and Disallowed 0.5%.

Size-exclusion chromatography. Size-exclusion chromatography runs were carried out on a Superdex 200 HR 16/60 gel filtration column (GE Healthcare) and monitored using the ultraviolet absorbance at 280 nm. Each injection used 60 µg Xpot in a final volume of 1 ml. For the binary and ternary complexes, 1.25 mg RanGTP and 750 µg yeast Phe-tRNA (Sigma) were used. Samples from the peak fraction were separated on 4–12% gradient MOPS SDS–PAGE gels (Invitrogen) and visualized by silver-staining.

- Walker, P. A. *et al.* Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology* **12**, 601–605 (1994).
- Doublie, S. *et al.* Crystallization and preliminary X-ray analysis of the 9 kDa protein of the mouse signal recognition particle and the selenomethionyl-SRP9. *FEBS Lett.* **384**, 219–221 (1996).
- Leslie, A. G. W. in *Joint CCP4 + ESRF-EAMCB Newsletter on Protein Crystallography* No. 26 (CCP4, 1992).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Broennimann, Ch *et al.* The PILATUS 1M detector. *J. Synchrotron Radiat.* **13**, 120–130 (2006).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
- de La Fortelle, E. & Bricogne, G. in *Macromolecular Crystallography Part A* (eds Sweet, R. M. & Carter, C. W. J.) 472–494 (Academic, 1997).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).
- Jones, T. A., Bergdoll, M. & Kjeldgaard, M. in *Crystallographic and Modeling Methods in Molecular Design* (eds Bugg, C. & Ealick, S.) 189–195 (Springer-Verlag, 1990).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
- Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).

LETTERS

The remnants of galaxy formation from a panoramic survey of the region around M31

Alan W. McConnachie¹, Michael J. Irwin², Rodrigo A. Ibata³, John Dubinski⁴, Lawrence M. Widrow⁵, Nicolas F. Martin⁶, Patrick Côté¹, Aaron L. Dotter⁷, Julio F. Navarro⁷, Annette M. N. Ferguson⁸, Thomas H. Puzia¹, Geraint F. Lewis⁹, Arif Babul⁷, Pauline Barmby¹⁰, Olivier Bienaymé³, Scott C. Chapman², Robert Cockcroft¹¹, Michelle L. M. Collins², Mark A. Fardal¹², William E. Harris¹¹, Avon Huxor¹³, A. Dougal Mackey⁸, Jorge Peñarrubia², R. Michael Rich¹⁴, Harvey B. Richer¹⁵, Arnaud Siebert³, Nial Tanvir¹⁶, David Valls-Gabaud¹⁷ & Kimberly A. Venn⁷

In hierarchical cosmological models¹, galaxies grow in mass through the continual accretion of smaller ones. The tidal disruption of these systems is expected to result in loosely bound stars surrounding the galaxy, at distances that reach 10–100 times the radius of the central disk^{2,3}. The number, luminosity and morphology of the relics of this process provide significant clues to galaxy formation history⁴, but obtaining a comprehensive survey of these components is difficult because of their intrinsic faintness and vast extent. Here we report a panoramic survey of the Andromeda galaxy (M31). We detect stars and coherent structures that are almost certainly remnants of dwarf galaxies destroyed by the tidal field of M31. An improved census of their surviving counterparts implies that three-quarters of M31's satellites brighter than $M_v = -6$ await discovery. The brightest companion, Triangulum (M33), is surrounded by a stellar structure that provides persuasive evidence for a recent encounter with M31. This panorama of galaxy structure directly confirms the basic tenets of the hierarchical galaxy formation model and reveals the shared history of M31 and M33 in the unceasing build-up of galaxies.

Precise measurements of stars and star clusters within the Milky Way have contributed significantly to the development of a cosmological model for galaxy formation^{5,6}. The discovery that the Sagittarius dwarf galaxy was being cannibalized by the Milky Way⁷ brought into sharp focus the role of satellite accretion in the build-up of a galaxy's mass. Models now propose that galaxies form within dark-matter haloes that grow through the continual accretion and merger of smaller sub-haloes. The predicted number of sub-haloes is at least a few orders of magnitude more than the number of dwarf galaxies observed as satellites around the Milky Way^{8,9}. If we assume that the underlying cosmology is correct, this implies that either significant numbers of satellites remain undiscovered or only a fraction of sub-haloes contain baryons (stars and gas). In either case, the number, luminosity and spatial distributions of satellite galaxies are important but poorly determined quantities whose values depend strongly on the processes through which baryons are accreted and retained by sub-haloes^{10,11}. Many of the luminous sub-haloes are expected to be perturbed, and even shredded, by the tidal field of

the host galaxy, leaving behind stellar debris in the form of streams and substructures within a diffuse stellar halo^{2,3}.

Systematic studies of the Milky Way's stellar halo, such as with the Sloan Digital Sky Survey¹², have recently revealed a large number of dwarf galaxies and tidal streams. However, our viewpoint from within the Milky Way introduces selection effects and difficulties in interpretation, making a homogeneous global study difficult. To this end, we have initiated the 'Pan-Andromeda Archaeological Survey' (PAndAS), a programme using the 1-square-degree field-of-view MegaPrime/MegaCam camera on the 3.6-m Canada–France–Hawaii Telescope (CFHT). We are imaging the closest spiral galaxy, M31, and its less massive companion M33. Once completed, the survey will cover more than 300 square degrees (more than 70,000 kpc²) and extend to a maximum projected radius from M31's centre of $r_p \approx 150$ kpc. It is the largest contiguous imaging survey of a massive galaxy and spans the stellar halo out to extremely large radii. PAndAS builds on earlier Isaac Newton Telescope and CFHT photometric studies of this galaxy^{13–17}. It surveys in the g (4,140–5,600 Å) and i (7,020–8,530 Å) bands and resolves individual stars in M31 to depths of $g = 26.5$, $i = 25.5$ at a signal-to-noise ratio of 10. The programme started in August 2008 and will continue until January 2011.

Figure 1 shows the spatial density distribution of sources in our extant PAndAS fields that are consistent with red giant branch (RGB) stars in M31. Although the region surveyed so far covers ~ 220 square degrees (nearly 100-fold the area of the classical optical disk of M31), we find RGB stars everywhere across our survey. These stars trace the low-luminosity structure of the galaxy and reveal the vast extent of M31, challenging the commonly held impression of the size of this typical bright galaxy (and, by extension, other galaxies of similar luminosities). These stars are unlikely to have formed *in situ* at these radii because it is improbable that the local density of gas was high enough to promote star formation. Instead, it is expected that the stars have been accreted from dwarf galaxies or proto-galactic fragments, a conclusion that is consistent with the basic tenets of hierarchical galaxy formation.

Our interpretation of halo stars as accreted relics is supported by the presence of multiple, large, coherent substructures over our survey

¹NRC Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, British Columbia, Canada V9E 2E7. ²Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. ³Observatoire de Strasbourg, 11, rue de l'Université, F-67000 Strasbourg, France. ⁴Department of Astronomy & Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario, Canada M5S 3H4. ⁵Department of Physics, Engineering Physics, and Astronomy Queen's University, Kingston, Ontario, Canada K7L 3N6. ⁶Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. ⁷Department of Physics and Astronomy, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, Canada V8P 5C2. ⁸Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK. ⁹Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia. ¹⁰Department of Physics and Astronomy, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada N6A 3K7. ¹¹Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada L8S 4M1. ¹²University of Massachusetts, Department of Astronomy, LGRT 619-E, 710 N. Pleasant Street, Amherst, Massachusetts 01003-9305, USA. ¹³Department of Physics (Astrophysics Group), H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK. ¹⁴Department of Physics and Astronomy, University of California, Los Angeles, PAB, 430 Portola Plaza, Los Angeles, California 90095-1547, USA. ¹⁵Department of Physics and Astronomy, 6224 Agricultural Road, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1. ¹⁶Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ¹⁷Laboratoire Galaxies et Étoiles, Physique et Instrumentation, CNRS UMR 8111, Observatoire de Paris, 5 Place Jules Janssen, 92195 Meudon, France.

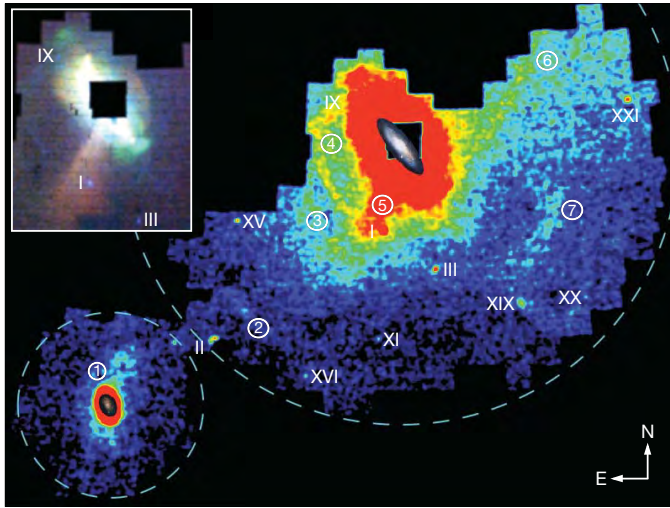


Figure 1 | Stellar density map of Andromeda–Triangulum. A tangent-plane projection of the density distribution of stellar sources in our extant PAndAS imaging is shown, with colours and magnitudes consistent with RGB stars at the distance of M31. The inset shows the central parts of our survey at higher resolution. Dashed circles represent the maximum projected radii of 150 and 50 kpc from M31 and M33, respectively. Scale images of the disks of M31 and M33 are overlaid. Visible dwarf satellites are indicated with roman numerals. Numbers in circles indicate the largest and most obvious substructures detected in a visual inspection of the image: 1, M33 structure; 2, 125-kpc stream (stream A); 3, stream C; 4, eastern arc (stream D); 5, giant stellar stream; 6, northwest minor-axis stream; 7, southwest cloud. Features 1, 6 and 7 (and part of 4) are new discoveries. Stellar sources were identified by using star–galaxy classification techniques described previously^{14,16}. Candidate RGB stars were selected by their position in a colour–magnitude diagram relative to theoretical isochrones²⁹ for a 12-Gyr population at the distance of M31 (ref. 26), using only stellar sources with $i_0 < 23.5$. A projection of the stellar density distribution within a putative metallicity range $-2.5 < [\text{Fe}/\text{H}] < -0.6$ dex was created with $0.02^\circ \times 0.02^\circ$ pixels, smoothed with a Gaussian filter with dispersion 3 arcmin, and displayed with square-root scaling. The inset was created by combining, with a red–green–blue colour scheme, three such maps of M31 with metallicities of $-0.4 < [\text{Fe}/\text{H}] < +0.2$ dex, $-1.3 < [\text{Fe}/\text{H}] < -0.4$ dex and $-2.3 < [\text{Fe}/\text{H}] < -1.3$ dex, respectively (each with $0.01^\circ \times 0.01^\circ$ pixels, smoothed with a Gaussian filter with 2 arcmin dispersion) and displayed with logarithmic scaling. Not all structures are visible with this (or any other) choice of metallicity cut, filter and scaling.

area. Some of these features, such as the giant stellar stream¹³ (no. 5 in Fig. 1), were previously known. New structures discovered in our survey include a radial overdensity along the northwest minor axis extending nearly 100 kpc from M31 (no. 6), a diffuse structure to the southwest coherent over an arc spanning ~ 40 kpc at a distance of ~ 100 kpc from M31 (no. 7), and an apparent continuation of a previously known stream in the east¹⁶, ~ 50 kpc from M31 that loops around to the north of the galaxy (no. 4). The large scale of the new structures is striking, as is their distance from the centre of M31. They are expected to maintain coherence for at least a few gigayears in the inner halo ($r_p < 50$ kpc), and at least a Hubble time in the outer halo ($r_p > 100$ kpc)¹⁸. The implication is that these structures are the remains of previously accreted dwarf galaxies. None of the newly discovered stellar structures clearly correlate with H I detections in the environs of M31 (ref. 19). Although the lack of young stars tracing sites of star formation implies that high H I column densities are not expected, more detailed predictions on the H I content of these substructures are not yet possible.

The surviving counterparts of the progenitors of these substructures—the dwarf galaxies—are visible in Fig. 1 as concentrated, round, overdensities, and the large area surveyed permits an improved census of these objects. In Fig. 2a we plot the projected radial number density distribution of galaxies around M31; the

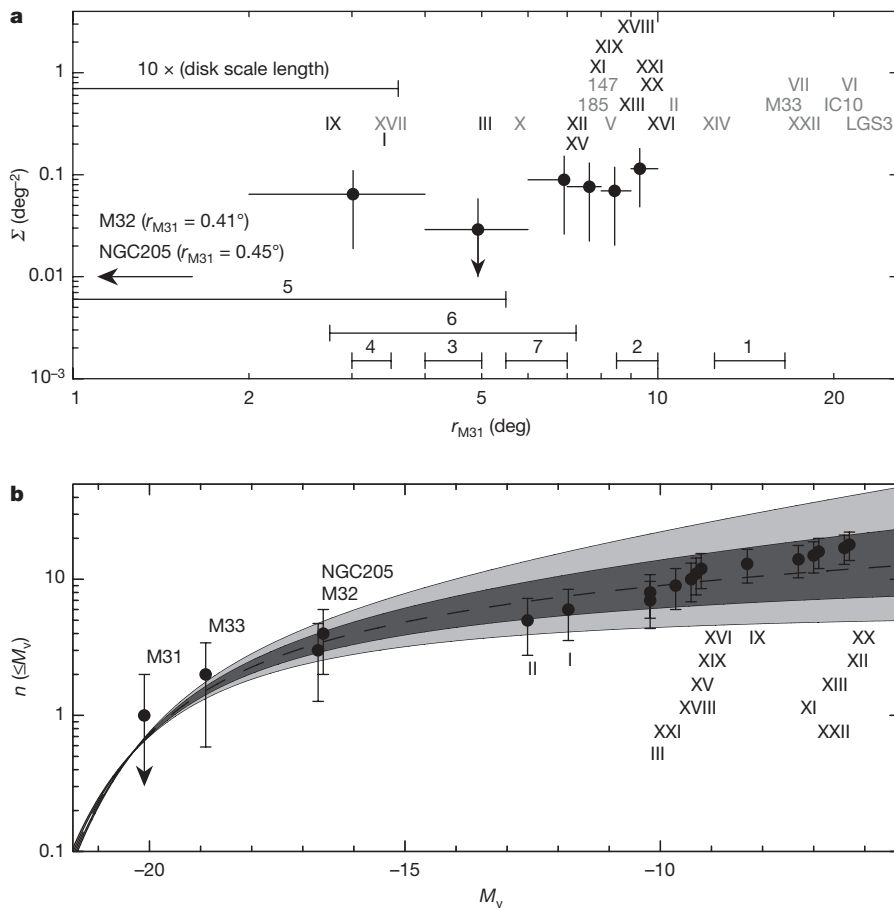
profile shows no sign of declining within 150 kpc. In Fig. 2b we plot the cumulative luminosity distribution of all galaxies around M31. We find that it is well described by the usual Schechter function²⁰ with a faint-end slope of $\alpha = -0.98 \pm 0.07$, where the uncertainty represents the standard deviation. These distributions imply that 6 ± 4 satellites more luminous than $M_v \approx -6$ still remain to be discovered within 150 kpc (10°) of M31. Extrapolating the observed flat number-density profile outwards then suggests that M31 may have as many as 88 ± 20 such satellites out to $r_p \approx 300$ kpc (roughly equal to the expected virial radius of M31's dark-matter halo; see Supplementary Information), only about one-quarter of which are currently known. The derived number is an upper limit, because we expect the radial profile to decline beyond the survey area. Thus, even accounting for observational incompleteness, there is still an order of magnitude too few satellites compared with the expected number of dark-matter haloes. If some of these haloes do not contain stars, then a comprehensive inventory of those that do may shed much light on the solution to the 'missing satellites problem'.

The brightest of M31's satellite companions, M33, is surrounded by a previously unknown prominent stellar structure (no. 1 in Fig. 1). This feature has an extension stretching $\sim 2^\circ$ (~ 30 kpc projected) to the northwest towards M31, nearly three times farther out than the classical disk of M33. A second extension is also visible in the south. One possible origin of this structure is that it is the accreted remains of a dwarf satellite, similar to the structures observed around M31. However, a long-standing puzzle about M33 is the existence of an extremely warped H I disk, with no apparent counterpart in the 'pristine' stellar disk^{21,22}. What process could distort the H I disk but leave the stellar disk unaffected? The newly discovered feature has a similar orientation to the H I warp, suggesting that we have discovered its optical counterpart. Furthermore, the general northwest–southeast symmetry of the distortion, and the broad alignment of the northwest extension with the measured transverse velocity of M33 (ref. 23), are evidence of a tidal disturbance excited as this galaxy orbits around M31.

We test this hypothesis by considering the M33–M31 orbit. If the orbit of M33 carries it close to M31, the tidal field of the latter will excite the M33 disk and eject stars, but the orbit cannot be so close that M33 is severely distorted or disrupted²⁴. We use the techniques described in the Supplementary Information to search for M33 orbits that are consistent with the known constraints and have reasonably close encounters with M31. The search reveals that the smallest possible pericentre distance is ~ 40 kpc, with close passage occurring a few billion years ago. We select several representative orbits and carry out n -body simulations²⁵ of the encounter.

Figure 3 and Supplementary Movie 1 show results from one plausible interaction model. This orbit reproduces with good accuracy the observed distances²⁶, angular positions and radial velocities of M31 and M33, as well as the proper motion of M33 (ref. 23). The interaction excites tidal tails in the M33 disk that wind up to form an extended distribution of stars with dimensions similar to those of the debris observed in the PAndAS data. In addition, the debris is warped away from the disk in the same sense as the known gaseous warp^{21,22}, lending further support to the interaction hypothesis.

Finally, we note that the encounter between M31 and M33 leaves an imprint on the larger system, exciting a mild warp and disturbances in the M31 disk at large radii. These phenomena are consistent with some of the unusual features observed in the outer regions of M31 (ref. 14), particularly with the discovery of younger stellar populations in these fields²⁷ and with the measured rotational signature shared by many of the substructures²⁸. It is plausible that these stars were originally formed in the thin disk and excited to their present locations by a galactic interaction, perhaps with M33 at an earlier phase in its orbit. The unrivalled panorama of galaxy structure presented here for M31 therefore reveals the continuing role of accretion and interactions in shaping its properties, and is a startling visual demonstration of the truly vast scale of galaxies.

**Figure 2 | Distribution of M31 dwarf galaxies.**

a, Points show the projected radial number density of dwarfs (Σ), derived for galaxies within 10° of M31 in the survey region. Horizontal error bars show the size of the bin, with points plotted at the mean galaxy radius in that bin. Vertical error bars are Poissonian. Galaxy names in black indicate the positions of the galaxies contributing to each bin; the remaining galaxies are named in grey. The radial ranges of the substructures indicated in Fig. 1 are also shown. **b**, The luminosity distribution, using all galaxies within the survey. Error bars are Poissonian. The dashed line shows the best-fit Schechter function, with slope $\alpha = -0.98 \pm 0.07$. Dark and light grey areas show 1σ and 2σ deviations in α , respectively. In both panels we are probably missing some satellites within $\sim 4^\circ$ of M31, where the high stellar density makes it difficult to detect faint satellites. We can detect galaxies brighter than $M_v \approx -6$, although derivation of the exact incompleteness levels remains to be calculated for the completed survey. To avoid these selection effects, the slope of the luminosity function is derived by using a χ^2 fit of a Schechter function to galaxies brighter than $M_v < -8$, where incompleteness is not an issue, and we note that the extrapolation to lower luminosities fits these data well. Because of the lack of bright galaxies, we cannot independently derive M^* . Instead, we fix M^* at a range of values and derive corresponding values of α , finding that α is robust to within 1σ for $-19.5 > M^* > -21.5$ (with a trend such that fainter values for M^* correspond to lower values for α). The quoted value of α corresponds to $M^* = -21$.

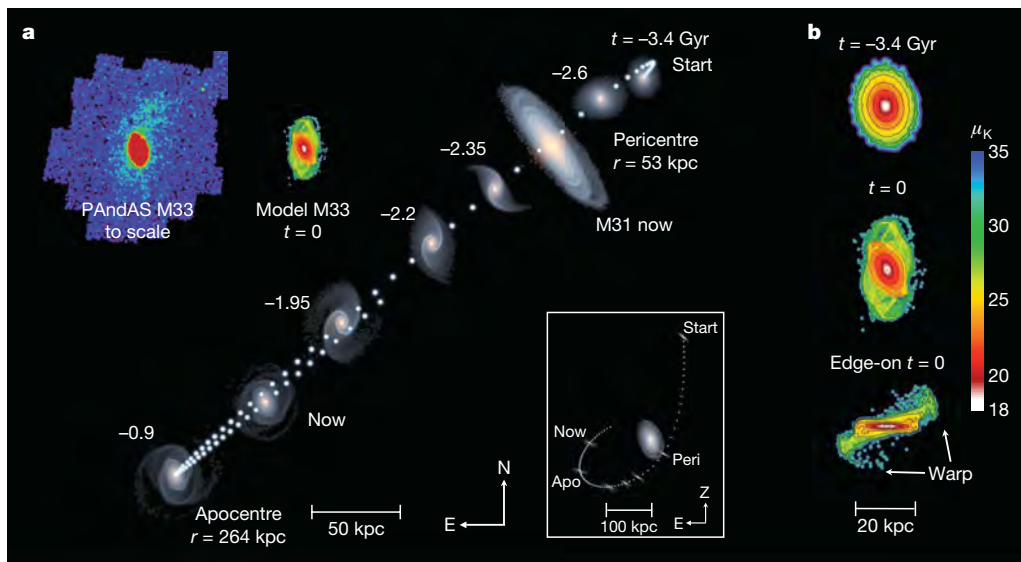


Figure 3 | An M31-M33 interaction. **a**, The evolution of M33 about M31 along an orbit consistent with the angular positions, distances²⁶ and radial velocities of M31 and M33 and with M33's proper motion²³. Here, M33 starts 3.4 Gyr ago at a distance of $r \approx 200$ kpc on the far side of M31 falling down the line of sight (Z) to the Milky Way. After 800 Myr, M33 reaches pericentre and proceeds across our line of sight towards the southeast, reaching apocentre about 900 Myr ago before falling back towards M31 to its current position. Dots tracing the orbit are separated by 49 Myr to give a sense of the speed along the orbit. The lower inset shows a perpendicular view of M33's orbit. **b**, Quantification of the expected K-band surface brightness of M33 at

different times from face-on and edge-on perspectives. The inner red/orange region with $\mu_K < 22$ mag arcsec $^{-2}$ defines the size of the usual optical disk of M33 seen in images. The initial equilibrium models³⁰ for the two galaxies consist of a disk, bulge and dark-matter halo with structural parameters that accurately reproduce the observed surface brightness profiles and rotation curves. Because the mass profile of M31 beyond 100 kpc is not well constrained by observations, we appeal to cosmological arguments²⁴ that predict a mass of 2.5×10^{12} solar masses within $r < 280$ kpc. See the Supplementary Information for more details.

Received 11 June; accepted 23 July 2009.

1. White, S. D. M. & Rees, M. J. Core condensation in heavy halos—a two-stage theory for galaxy formation and clustering. *Mon. Not. R. Astron. Soc.* **183**, 341–358 (1978).
2. Bullock, J. S. & Johnston, K. V. Tracing galaxy formation with stellar halos. I. Methods. *Astrophys. J.* **635**, 931–949 (2005).
3. Abadi, M. G., Navarro, J. F. & Steinmetz, M. Stars beyond galaxies: the origin of extended luminous haloes around galaxies. *Mon. Not. R. Astron. Soc.* **365**, 747–758 (2006).
4. Johnston, K. V. *et al.* Tracing galaxy formation with stellar halos. II. Relating substructure in phase and abundance space to accretion histories. *Astrophys. J.* **689**, 936–957 (2008).
5. Eggen, O. J., Lynden-Bell, D. & Sandage, A. R. Evidence from the motions of old stars that the Galaxy collapsed. *Astrophys. J.* **136**, 748–766 (1962).
6. Searle, L. & Zinn, R. Compositions of halo clusters and the formation of the galactic halo. *Astrophys. J.* **225**, 357–379 (1978).
7. Ibata, R. A., Gilmore, G. & Irwin, M. J. A dwarf satellite galaxy in Sagittarius. *Nature* **370**, 194–196 (1994).
8. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where are the missing galactic satellites? *Astrophys. J.* **522**, 82–92 (1999).
9. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
10. Bullock, J. S., Kravtsov, A. V. & Weinberg, D. H. Reionization and the abundance of galactic satellites. *Astrophys. J.* **539**, 517–521 (2000).
11. Kravtsov, A. V., Gnedin, O. Y. & Klypin, A. A. The tumultuous lives of galactic dwarfs and the missing satellites problem. *Astrophys. J.* **609**, 482–497 (2004).
12. Belokurov, V. *et al.* The field of streams: Sagittarius and its siblings. *Astrophys. J.* **642**, L137–L140 (2006).
13. Ibata, R., Irwin, M., Lewis, G., Ferguson, A. M. N. & Tanvir, N. A giant stream of metal-rich stars in the halo of the galaxy M31. *Nature* **412**, 49–52 (2001).
14. Ferguson, A. M. N., Irwin, M. J., Ibata, R. A., Lewis, G. F. & Tanvir, N. R. Evidence for stellar substructure in the halo and outer disk of M31. *Astron. J.* **124**, 1452–1463 (2002).
15. Martin, N. F. *et al.* Discovery and analysis of three faint dwarf galaxies and a globular cluster in the outer halo of the Andromeda galaxy. *Mon. Not. R. Astron. Soc.* **371**, 1983–1991 (2006).
16. Ibata, R. *et al.* The haunted halos of Andromeda and Triangulum: a panorama of galaxy formation in action. *Astrophys. J.* **671**, 1591–1623 (2007).
17. McConnachie, A. W. *et al.* A trio of new Local Group galaxies with extreme properties. *Astrophys. J.* **688**, 1009–1020 (2008).
18. Johnston, K. V., Hernquist, L. & Bolte, M. Fossil signatures of ancient accretion events in the halo. *Astrophys. J.* **465**, 278–287 (1996).
19. Thilker, D. A. *et al.* On the continuing formation of the Andromeda galaxy: detection of H I clouds in the M31 halo. *Astrophys. J.* **601**, L39–L42 (2004).
20. Schechter, P. An analytic expression for the luminosity function for galaxies. *Astrophys. J.* **203**, 297–306 (1976).
21. Rogstad, D. H., Wright, M. C. H. & Lockhart, I. A. Aperture synthesis of neutral hydrogen in the galaxy M33. *Astrophys. J.* **204**, 703–711 (1976).
22. Corbelli, E. & Schneider, S. E. A warped disk model for M33 and the 21 centimeter line width in spiral galaxies. *Astrophys. J.* **479**, 244–257 (1997).
23. Brunthaler, A., Reid, M. J., Falcke, H., Greenhill, L. J. & Henkel, C. The Geometric distance and proper motion of the Triangulum Galaxy (M33). *Science* **307**, 1440–1443 (2005).
24. Loeb, A., Reid, M. J., Brunthaler, A. & Falcke, H. Constraints on the proper motion of the Andromeda galaxy based on the survival of its satellite M33. *Astrophys. J.* **633**, 894–898 (2005).
25. Dubinski, J. A parallel tree code. *N. Astron.* **1**, 133–147 (1996).
26. McConnachie, A. W. *et al.* Distances and metallicities for 17 Local Group galaxies. *Mon. Not. R. Astron. Soc.* **356**, 979–997 (2005).
27. Richardson, J. C. *et al.* The nature and origin of substructure in the outskirts of M31. I. Surveying the stellar content with the *Hubble Space Telescope* Advanced Camera for Surveys. *Astron. J.* **135**, 1998–2012 (2008).
28. Ibata, R. *et al.* On the accretion origin of a vast extended stellar disk around the Andromeda galaxy. *Astrophys. J.* **634**, 287–313 (2005).
29. Dotter, A. *et al.* The Dartmouth stellar evolution database. *Astrophys. J. Suppl. Ser.* **178**, 89–101 (2008).
30. Widrow, L. M., Pym, B. & Dubinski, J. Dynamical blueprints for galaxies. *Astrophys. J.* **679**, 1239–1259 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the entire staff at the Canada–France–Hawaii Telescope for taking the data, for initial processing with Elixir and for their continuing support throughout this project. A.M.N.F. and A.D.M. are supported by a Marie Curie Excellence Grant from the European Commission under contract MCEXT-CT-2005-025869. G.F.L. thanks the Australian Nuclear Science and Technology Organisation (ANSTO) for supporting his involvement in PAndAS through its Access to Major Research Facilities Program (AMRFP). R.M.R. acknowledges grant AST-0709479 from the National Science Foundation, and grants GO-9453, GO-10265 and GO-10816 from the Space Telescope Science Institute. The image of M33 overlaid in Fig. 1 is reproduced by courtesy of T. A. Rector and M. Hanna.

Author Contributions All authors assisted in the development and writing of the paper. In addition, A.W.M. is the Principal Investigator of PAndAS; M.J.I. led the data processing effort; R.A.I. was the Principal Investigator of an earlier CFHT MegaPrime/MegaCam survey, which PAndAS builds on (which included S.C.C., A.M.N.F., M.J.I., G.F.L., N.F.M., A.W.M. and N.T.); J.D., L.M.W. modelled the M31–M33 interaction; N.F.M. had a lead role in the study of the dwarf galaxies; P.C. assisted with constructing the luminosity function; and A.L.D. developed the theoretical isochrones.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.W.M. (alan.mcconnachie@nrc-cnrc.gc.ca).

LETTERS

Laser cooling by collisional redistribution of radiation

Ulrich Vogl¹ & Martin Weitz¹

The general idea that optical radiation may cool matter was put forward 80 years ago¹. Doppler cooling of dilute atomic gases is an extremely successful application of this concept^{2,3}. More recently, anti-Stokes cooling in multilevel systems has been explored^{4,5}, culminating in the optical refrigeration of solids^{6–9}. Collisional redistribution of radiation has been proposed¹⁰ as a different cooling mechanism for atomic two-level systems, although experimental investigations using moderate-density gases have not reached the cooling regime¹¹. Here we experimentally demonstrate laser cooling of an atomic gas based on collisional redistribution of radiation, using rubidium atoms in argon buffer gas at a pressure of 230 bar. The frequent collisions in the ultradense gas transiently shift a highly red-detuned laser beam (that is, one detuned to a much lower frequency) into resonance, whereas spontaneous decay occurs close to the unperturbed atomic resonance frequency. During each excitation cycle, kinetic energy of order $k_B T$ —that is, the thermal energy (k_B , Boltzmann's constant; T , temperature)—is extracted from the dense atomic sample. In a proof-of-principle experiment with a thermally non-isolated sample, we demonstrate relative cooling by 66 K. The cooled gas has a density more than ten orders of magnitude greater than the typical values used in Doppler-cooling experiments, and the cooling power reaches 87 mW. Future applications of the technique may include supercooling beyond the homogeneous nucleation temperature^{12,13} and optical chillers⁹.

Collisional redistribution is perhaps most widely known in the context of magneto-optical trapping of ultracold atoms, where it is a primary cause of trap loss processes¹⁴. In the long-studied field of room-temperature interatomic collisions, redistribution of radiation is a natural consequence of line-broadening effects due to collisionally aided excitation^{15–18}. A noteworthy issue is the extreme elasticity of collisions of excited-state alkali atoms with atomic noble buffer gases¹⁹. We have recently shown that under conditions of ultrahigh buffer-gas pressure, the frequent collisions allow for thermal equilibrium of coupled atom–light states²⁰, which may allow for a possible Bose–Einstein-like phase transition of atom–light quasiparticles^{21–23}.

The principle of the cooling effect we report here is illustrated in Fig. 1a. We use rubidium atoms in buffer gas at a pressure of several hundred bars, for which the collisionally broadened linewidth approaches the thermal energy, $k_B T$, in frequency units. In a quasi-molecular binary-collision picture, the energy splitting between the rubidium 5s ground state and the 5p excited state varies critically with the instantaneous spatial distance between the rubidium atom and a noble-gas perturber. For a red laser detuning of a few nanometres, that is, one far outside the usual rubidium thin-vapour absorption profile, excitation is possible when a noble-gas perturber approaches the rubidium atom and transiently shifts the D lines into resonance by means of the intermolecular potential; this is an example of collisionally aided excitation¹⁷. The 27-ns natural lifetime of the rubidium 5p state is orders of magnitude longer than the approximately picosecond duration of a collisional process. Provided that the collisions are elastic and the mean interatomic spacing is larger than the range of the potential, spontaneous decay of the excited state will most likely occur when the atom–perturber distance is greater than the range of the interatomic potential. The frequency of the emitted photon will correspondingly be closer to the unshifted atomic transition frequency. This completes the cooling cycle.

For red laser detuning, in each excitation cycle kinetic energy is extracted from the sample. Moreover, entropy can be transferred to the photonic field, as the narrowband exciting laser radiation is converted into scattered fluorescence with frequency bandwidth distributed over the pressure-broadened linewidth. The typical energy that can be removed from the sample in a cooling cycle is of the order of the thermal energy, $k_B T$, corresponding to $\sim h \times 10^{13}$ Hz (h , Planck's constant) for temperatures, T , of order of the room temperature. This is at least four orders of magnitude greater than the energy difference removed by a fluorescence photon in ordinary Doppler cooling of dilute atomic gases. We are aware that the described simple picture neglects multiparticle collisions.

To demonstrate the principle of laser cooling by collisional redistribution, we use a rubidium cell at an initial temperature of

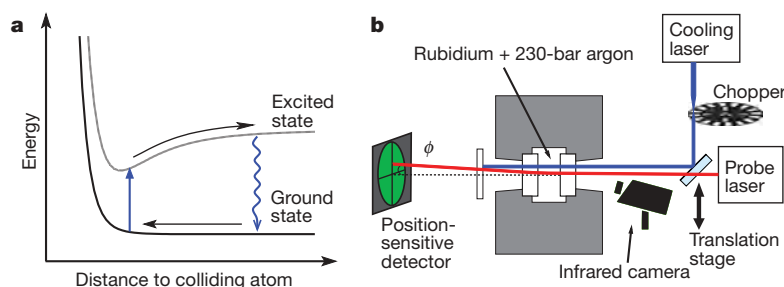


Figure 1 | Cooling principle and set-up. **a**, The black lines indicate how rubidium atomic energy levels change as functions of the separation from a noble-gas perturber (see ref. 30 for more realistic, modelled potentials for alkali noble gases). Highly red-detuned laser radiation can be absorbed when a perturber atom approaches the rubidium atom. Spontaneous decay occurs

mostly at larger perturber distances, so the emitted photons have higher frequencies than the incident radiation and the atomic ensemble loses energy. **b**, Experimental set-up for laser cooling of an ultradense gas by collisional redistribution of fluorescence.

¹Institut für Angewandte Physik der Universität Bonn, Wegelerstraße 8, 53115 Bonn, Germany.

620 K (vapour-limited number density, $n_{\text{Rb}} \approx 10^{16} \text{ cm}^{-3}$) and argon buffer gas at a pressure of 230 bar (23 MPa; $n_{\text{Ar}} \approx 10^{21} \text{ cm}^{-3}$). Tunable laser radiation is sent into the steel high-pressure gas cell through sapphire optical windows. The experimental set-up is shown in Fig. 1b and described in more detail in Methods Summary.

In initial experiments, the average optical-frequency difference between incident and fluorescence light was measured. From energy conservation, the expected laser cooling power in a sample is

$$P_{\text{cool,fl}} = P_{\text{opt}} a(\nu) \frac{\nu_{\text{fl}} - \nu}{\nu} \quad (1)$$

where P_{opt} denotes optical-beam power, $a(\nu)$ is the absorption probability for an incident optical frequency ν and ν_{fl} is the average frequency of the scattered fluorescence. Figure 2a shows typical results for the spectral distribution of the fluorescence for one red detuning (red line) and one blue detuning (a detuning to higher frequencies; blue line) relative to the rubidium D lines. The corresponding incident optical frequencies are indicated by dashed vertical lines. In both cases, the fluorescence spectrum is efficiently redistributed towards the line centre. The average frequency shift here reaches 15–20 THz. The corresponding differential energy will cool the atomic sample for red laser detuning and heat it for blue detuning. The connected squares in Fig. 2b show the expected cooling power as a function of the laser frequency, as derived by recording a number of different

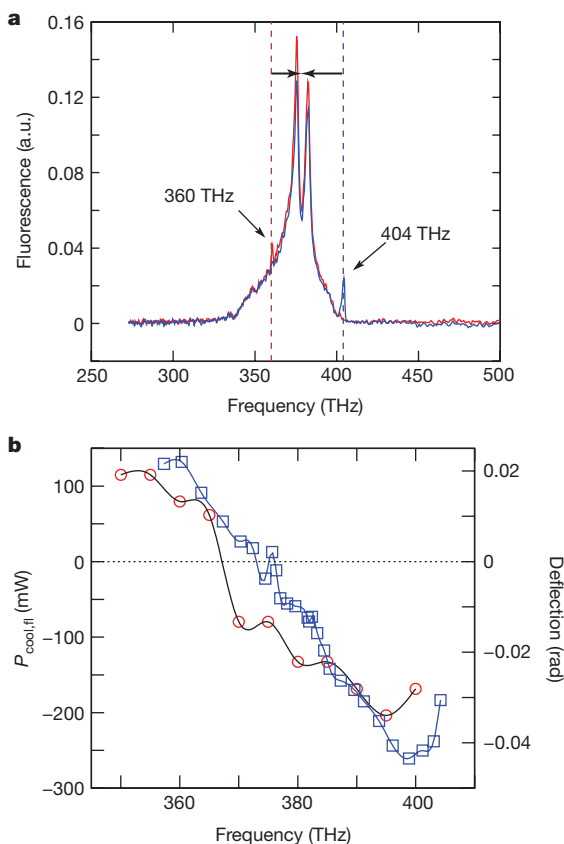


Figure 2 | Experimental spectra. **a**, Fluorescence spectra of the rubidium D lines at argon buffer-gas pressure of 230 bar for incident-laser frequencies of 360 THz (red line) and 404 THz (blue line), corresponding to red and blue detunings, respectively. The incident-laser frequencies are indicated as dashed vertical lines. In both cases, the fluorescence frequency is efficiently redistributed towards the centre of the D lines, as indicated. The visible small spikes are attributed mainly to stray excitation light. The rubidium resonance frequencies are 377 THz and 384 THz for the D1 and D2 lines, respectively. a.u., arbitrary units. **b**, The connected blue squares show the expected cooling power (left-hand axis) as a function of laser frequency, as derived from the average fluorescence frequency shift. The red circles (fitted with a spline function) show the angular deflection (right-hand axis) of a probe beam.

spectra such as that in Fig. 2a and determining the cooling power, following equation (1), from the average shift in the scattered fluorescence and the laser power absorbed in the rubidium sample. It is clear that energy is extracted from the sample on the red side of the spectrum.

The first evidence for an actual cooling of the gas was obtained using an infrared camera sensitive to black-body radiation in the wavelength range of 7.5–13 μm , which we directed at the outer surface of a sapphire window of our pressure cell. In this mid-infrared spectral region, the sapphire window material is highly opaque and we expect the window material, unlike the gas sample, to emit black-body radiation characteristic of its temperature. Owing to thermal transport in the sapphire material, a cooling of the dense gas will cause a temperature variation also at the outer surface of the cell window. For this measurement, we determined the temperature change of the sapphire window following a 30-s optical-cooling period. Figure 3 shows the result recorded for a cooling-laser frequency of 365 THz, corresponding to a red detuning of 16 THz relative to the rubidium D-line centroid. The observed average temperature decrease is $0.31 \pm 0.03 \text{ K}$ near the beam centre. This result for the cooling of the outer surface of the cell window is limited mainly by heating of the highly thermally conducting sapphire window (thermal conductivity of sapphire, $\kappa_{\text{sapphire}} \approx 200 \text{ W cm}^{-1} \text{ K}^{-1}$) as a result of contact with the metal body of the cell, which for each absorbed photon experiences an energy increase equal to the photon energy. The temperature increase of the body of the metal cell is also clearly visible in the figure, and is attributed mainly to absorption of scattered fluorescence radiation. The observed temperature decrease of the cell window is in agreement with the result of a simple heat-transport model calculation. We interpret the observed change in black-body radiation as clear qualitative evidence for laser cooling based on collisional redistribution.

In subsequent experiments, we probed temperature information directly from inside the gas cell by means of thermal deflection spectroscopy^{24–26}. A (non-resonant) helium–neon laser probe beam is sent collinearly, and slightly displaced from the cooling beam, through the high-pressure cell, as indicated in Fig. 1b. Radial temperature variations induced by the cooling beam cause a gradient in the refractive index of the gas, and result in a deflection of the probe beam similar to that produced by a prism. The circles in Fig. 2b show the measured deflection angle as a function of cooling-laser frequency. For red detuning the probe beam is deflected towards the cooling beam, which indicates cooling, whereas for blue detuning

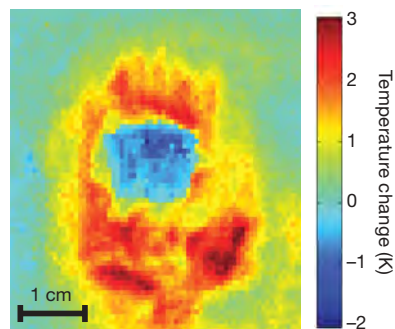


Figure 3 | Thermographic image. An infrared camera was directed at the sapphire entrance window of the high-pressure buffer-gas cell. Shown is the (average) colour-coded temperature change after a 30-s pulse of the cooling beam. The blue region in the centre is the sapphire window, cooled from behind by the gas sample. The surrounding metal flange appears red and yellow, owing to heating from absorption of scattered fluorescence radiation. To avoid cooling-laser radiation having a possible residual influence on the infrared-camera signal, we activated the camera only after the end of a cooling-laser pulse. The diameter of the cooling beam used in this measurement was 3 mm.

heating is observed. It is also apparent that the cooling regime is only reached when the laser frequency is detuned approximately 5 THz lower than the value at which it just equals the average fluorescence frequency; that is, at which the expected cooling power given by equation (1) reaches zero. This is attributed to residual heating, which could be due to a small quenching of excited-state rubidium atoms in the ultrahigh-pressure buffer-gas environment¹⁹. Our data could be explained by an average decrease in the 27-ns natural lifetime of the rubidium 5p state due to quenching in inelastic collisions by roughly 1 ns at an argon buffer-gas pressure of 230 bar, which would be consistent with the directly measured value of the lifetime change, 0 ± 2 ns. The ratio of elastic to inelastic collisions of excited-state rubidium atoms with argon rare-gas atoms here exceeds 20,000:1.

To acquire the spatial temperature profile induced by the cooling beam, we scanned the lateral offset between cooling and the probe beam. The obtained deflection data are shown by the red dots in Fig. 4 for a cooling-laser frequency of 365 THz and a relative absorption of 90% in the 1-cm-long gas cell, which fits well with results derived using a theoretical heat-transfer model (Fig. 4, black line). The cooling- and probe-beam diameters used in this measurement are 1 mm. We derived the lateral temperature profile near the cell entrance from numerical integration of the data (Methods Summary); the result is shown by the solid blue line in Fig. 4. The obtained temperature drop in the centre of the cooling beam is 66 ± 13 K. This value is much larger than the above-mentioned result for the temperature drop detected at the outer surface of the cell window. This is easily understood as resulting from the thermal conductivity of the argon gas being much less than that of the sapphire window material, meaning that the cooling within the gas volume near the focal region of the beam can proceed further.

The model we used to simulate heat transfer in the gas cell volume assumes that the cooling rate is proportional to the optical intensity of the Gaussian laser beam; that is, we neglect the influence of radiation transport, which assumes that $\omega \ll \ell_{\text{abs}}$, where ω denotes the cooling-beam radius and ℓ_{abs} is the absorption length. In general, we expect the influence of radiation transport to be weaker than it is in the area of ultracold atomic gases¹⁴, as the influence of light forces in the buffer-gas sample is much smaller, but we are aware that future work ought to investigate this effect in more detail. The dashed blue line in Fig. 4 is the temperature profile obtained from the theoretical heat-transfer model. A cooling power of 87 ± 20 mW can be estimated from the model, corresponding to a cooling efficiency of

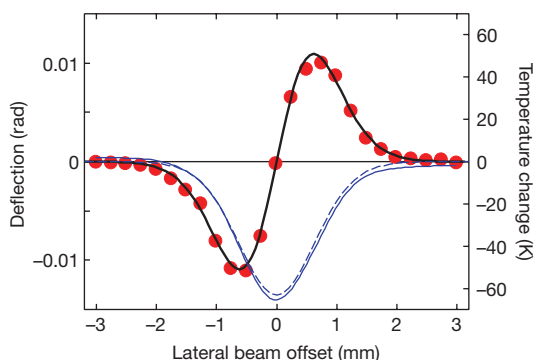


Figure 4 | Measurement of temperature inside the cell. The red dots show results for probe-beam angular deflection as a function of lateral offset between probe and cooling beams, and the black line gives a fit to a theoretical model. The deflection gives a measure of the radial temperature gradient induced by the cooling beam (Methods Summary). The solid blue line shows the corresponding temperature profile near the cell entrance. The determined temperature drop in the beam centre is 66 ± 13 K. The quoted error is determined mainly by uncertainties in the optical-beam diameter, the beam geometry, the measured absorption coefficient for the cooling radiation and the buffer-gas pressure measurement. The dashed blue line follows a heat-transport model (equation (2)).

$3.6 \pm 0.8\%$. These values are in reasonable agreement with the expected cooling power derived from the differential frequency shift of the fluorescence photons (see Fig. 2b and the corresponding discussion). We note that the cooling power is four to five orders of magnitude greater than that achieved in Doppler-cooling experiments and is also greater than results reported for the laser cooling of solids⁷.

In conclusion, evidence for redistribution laser cooling of a gas has been obtained from an observed blue shift of the scattered fluorescence, thermographic imaging of the cell window and thermal deflection spectroscopy. The measured temperature drop is presently limited by thermal conductivity of the argon gas, and could be improved either by focusing the cooling beam tightly into a relatively short, optically dense gas cell, to deposit the available cooling power into a smaller gas volume, or by the use of a thermally isolated all-sapphire cell.

METHODS SUMMARY

Set-up and preparatory measurements. We performed the experiment using a high-pressure stainless-steel cell that has a volume of a few cubic centimetres and sapphire optical windows and was filled with rubidium vapour and argon buffer gas²⁰. The optical cooling beam was derived from a continuous-wave titanium-sapphire laser with an output power of 3 W, and was directed into the pressure cell through the sapphire windows. To measure the rubidium excited-state lifetime in the pressure-broadened system, the laser radiation was periodically chopped using an acousto-optic modulator and the emitted fluorescence recorded with a photomultiplier tube.

Thermal deflection spectroscopy. To measure temperatures inside the cell, we periodically blocked the cooling beam with a mechanical shutter and monitored the deflection of a probe beam using a split photodiode. Temperature variations induced by the cooling beam cause a local density change and a variation of the refractive index, n , according to $dn/dT \approx -(n-1)/T$. A collinear probe beam displaced by a distance r is deflected by an angle of

$$\phi = \frac{n-1}{T} \int_0^L \frac{dT(r,z)}{dz} dz$$

where L denotes the cell length. Assuming that the cooling beam removes energy with a rate following its intensity distribution, $I(r,z) = I_0 \exp(-\alpha z - 2r^2/\omega^2)$, with absorption coefficient $\alpha = 1/\ell_{\text{abs}}$, for a long cell the heat transfer in the radial direction will dominate and the temperature distribution can be determined from the beam deflection as follows:

$$\Delta T(r,z) = \frac{T}{n-1} \frac{\alpha e^{-\alpha z}}{1 - e^{-\alpha L}} \int_r^\infty \phi(r') dr'$$

For the dense gas at the 632-nm probe-beam wavelength (which is far detuned from the rubidium D lines), we used a refractive index of $n \approx 1.057 \pm 0.006$ for argon²⁷; the uncertainty is predominantly determined by the accuracy of our cell pressure measurement.

For comparison, we fitted our deflection data assuming a temperature profile derived from a heat transport model^{25–27}

$$\Delta T_{\text{th}}(r,z) = \frac{P_{\text{cool}}}{4\pi\kappa} e^{-\alpha z} \left[\text{Ei} \left(\frac{-2r^2}{\omega^2 + 8Dt} \right) - \text{Ei} \left(\frac{-2r^2}{\omega^2} \right) \right] \quad (2)$$

where the cooling power, P_{cool} , and the thermal diffusivity, D , of the gas were left as free fit parameters; t denotes the exposure time; and a thermal conductivity of $\kappa = 0.034 \text{ W m}^{-1} \text{ K}^{-1}$ was used²⁸. At our pressure and temperature values, $D = \kappa/c\rho$, where c denotes the specific heat capacity and ρ is the density of the gas, is not very well known, but the derived value, $D = 0.16 \pm 0.05 \text{ mm}^2 \text{ s}^{-1}$, is in agreement with published results for argon^{28,29}.

Received 4 March; accepted 9 June 2009.

1. Pringsheim, P. Zwei Bemerkungen über den Unterschied von Lumineszenz- und Temperaturstrahlung. *Z. Phys.* **57**, 739–746 (1929).
2. Hänsch, T. W. & Schawlow, A. L. Cooling of gases by laser radiation. *Opt. Commun.* **13**, 68–69 (1975).
3. Chu, S., Cohen-Tanoudji, C. N. & Phillips, W. D. Nobel lectures in physics 1997. *Rev. Mod. Phys.* **70**, 685–741 (1998).
4. Djeu, N. & Whitney, W. T. Laser cooling by spontaneous anti-Stokes scattering. *Phys. Rev. Lett.* **46**, 236–239 (1981).

5. Zander, C. & Drexhage, K. H. in *Advances in Photochemistry* Vol. 20 (eds Neckers, D. C., Volman, D. H. & von Büna, G.) 59–78 (Wiley, 1995).
6. Epstein, R. I., Buchwald, M., Edwards, B., Gosnell, T. & Mungan, C. Observation of laser induced fluorescent cooling of a solid. *Nature* **377**, 500–503 (1995).
7. Hoyt, C. W. *et al.* Advances in laser cooling of thulium-doped glass. *J. Opt. Soc. Am. B* **20**, 1066–1074 (2003).
8. Thiede, J., Distel, J., Greenfield, S. R. & Epstein, R. I. Cooling to 208 K by optical refrigeration. *Appl. Phys. Lett.* **86**, 154107 (2005).
9. Sheik-Bahae, M. & Epstein, R. I. Optical refrigeration. *Nature Photon.* **12**, 693–699 (2007).
10. Berman, P. R. & Stenholm, S. Heating or cooling using collisionally aided fluorescence. *Opt. Commun.* **24**, 155–157 (1978).
11. Giacobino, E., Tawil, M., Berman, P. R., Redi, O. & Stroke, H. H. Production of “hot” excited-state atoms in collisionally aided radiative transitions. *Phys. Rev. A* **28**, 2555–2557 (1983).
12. Debenedetti, P. G. & Stanley, H. E. Supercooled and glassy water. *Phys. Today* **56**, 40–46 (2003).
13. Koop, T., Luo, B., Tsias, A. & Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **406**, 611–614 (2000).
14. Adams, C. S. & Riis, E. Laser cooling and trapping of neutral atoms. *Prog. Quantum Electron.* **21**, 1–79 (1997).
15. Cohen-Tannoudji, C., Dupont-Roc, J. & Grynberg, G. *Atom-Photon Interactions – Basic Processes and Applications* 490–514 (Wiley, 1992).
16. Schuller, F. & Behmenburg, W. Perturbation of spectral lines by atomic interactions. *Phys. Rep.* **12**, 273–334 (1974).
17. Yeh, S. & Berman, P. R. Theory of collisionally aided radiative excitation. *Phys. Rev. A* **19**, 1106–1116 (1979).
18. Hedges, R. E. M., Drummond, D. L. & Gallagher, A. Extreme-wing line broadening and Cs-inert-gas potentials. *Phys. Rev. A* **6**, 1519–1544 (1972).
19. Speller, E., Staudenmayer, B. & Kempter, V. Quenching cross sections for alkali-inert gas collisions. *Z. Phys. A* **291**, 311–318 (1979).
20. Vogl, U. & Weitz, M. Spectroscopy of atomic rubidium at 500-bar buffer gas pressure: approaching the thermal equilibrium of dressed atom-light states. *Phys. Rev. A* **78**, 011401 (2008).
21. Eastham, P. R. & Littlewood, P. B. Bose condensation of cavity polaritons beyond the linear regime: the thermal equilibrium of a model microcavity. *Phys. Rev. B* **64**, 235101 (2001).
22. Deng, H., Weihs, G., Santori, C., Bloch, J. & Yamamoto, Y. Condensation of semiconductor microcavity exciton polaritons. *Science* **298**, 199–202 (2002).
23. Bolkart, C., Weiss, R., Rostohar, D. & Weitz, M. Coherent and BCS-type quantum states of dark polaritons. *Laser Phys.* **15**, 3–6 (2005).
24. Whinnery, J. Laser measurement of optical absorption in liquids. *Acc. Chem. Res.* **7**, 225–231 (1974).
25. Jackson, W. B., Amer, N. M., Boccara, A. C. & Fournier, D. Photothermal deflection spectroscopy and detection. *Appl. Opt.* **20**, 1333–1344 (1981).
26. Spear, J. D., Russo, R. E. & Silva, R. J. Collinear photothermal deflection spectroscopy with light scattering samples. *Appl. Opt.* **29**, 4225–4234 (1990).
27. Born, M. & Wolf, E. *Principles of Optics* 7th edn, 89–115 (Pergamon, 1999).
28. Rosenbaum, B. M., Oshen, S. & Thodos, G. Thermal conductivity of argon in the dense gaseous and liquid regions. *J. Chem. Phys.* **44**, 2831–2838 (1966).
29. Tournier, J.-M. P. & El-Genk, M. S. Properties of noble gases and binary mixtures for closed Brayton cycle applications. *Energy Convers. Manage.* **49**, 469–492 (2008).
30. Pascale, J. & Vandeplanque, J. Excited molecular terms of the alkali-rare gas atom pairs. *J. Chem. Phys.* **60**, 2278–2289 (1974).

Acknowledgements We thank J. Nipper for experimental contributions during the early phase of this project. Financial support from the Deutsche Forschungsgemeinschaft within the focused research unit FOR557 is acknowledged.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.W. (martin.weitz@uni-bonn.de).

LETTERS

From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal

Jianping Zheng^{1*}, Jens J. Birktoft^{1*}, Yi Chen^{2*}, Tong Wang¹, Ruojie Sha¹, Pamela E. Constantinou^{1†}, Stephan L. Ginell³, Chengde Mao² & Nadrian C. Seeman¹

We live in a macroscopic three-dimensional (3D) world, but our best description of the structure of matter is at the atomic and molecular scale. Understanding the relationship between the two scales requires a bridge from the molecular world to the macroscopic world. Connecting these two domains with atomic precision is a central goal of the natural sciences, but it requires high spatial control of the 3D structure of matter¹. The simplest practical route to producing precisely designed 3D macroscopic objects is to form a crystalline arrangement by self-assembly, because such a periodic array has only conceptually simple requirements: a motif that has a robust 3D structure, dominant affinity interactions between parts of the motif when it self-associates, and predictable structures for these affinity interactions. Fulfilling these three criteria to produce a 3D periodic system is not easy, but should readily be achieved with well-structured branched DNA motifs tailed by sticky ends². Complementary sticky ends associate with each other preferentially and assume the well-known B-DNA structure when they do so³; the helically repeating nature of DNA facilitates the construction of a periodic array. It is essential that the directions of propagation associated with the sticky ends do not share the same plane, but extend to form a 3D arrangement of matter. Here we report the crystal structure at 4 Å resolution of a designed, self-assembled, 3D crystal based on the DNA tensegrity triangle⁴. The data demonstrate clearly that it is possible to design and self-assemble a well-ordered macromolecular 3D crystalline lattice with precise control.

The tensegrity triangle is a rigid DNA motif with three-fold rotational symmetry, consisting of three helices that are directed along linearly independent vectors, that is, their helix axis directions do not all share the same plane. The helices are connected pair-wise by three four-arm branched junctions so as to produce the stiff alternating over-and-under motif shown schematically in Fig. 1a. Thus, there are three helical domains, each containing two double helical turns (21 nucleotide pairs, including sticky ends). There are seven strands in the molecule, three that partake in a crossover near the corners (magenta in Fig. 1a), three that extend over each edge of the tensegrity triangle (green in Fig. 1a), and a final nicked strand at the centre (blue in Fig. 1a), completing the crossovers and the double helices between the crossovers; the green and magenta strands indicate an over-and-under motif. By tailing the three helices with short single-stranded cohesive segments ('sticky ends') the helices can be directed to connect with helices belonging to six other molecules in six different directions, thereby yielding a 3D periodic lattice, that is, a crystal.

The complementary GA and TC sequences of the sticky ends used here are indicated in red letters in Fig. 1a. In this case, we have worked with a three-fold symmetric system, because that design has produced the best crystals of this motif (~4 Å resolution). Thus, each

of the three magenta and three green strands contains the same sequence; the central 21-mer strand has a triply repeating sequence, and its nicked site is three-fold rotationally averaged, occurring with one-third occupancy in each edge; in this design, the six sticky ends form three identical complementary pairs. Triangles lacking this symmetry produced crystals that diffracted to a lower resolution. Indexing

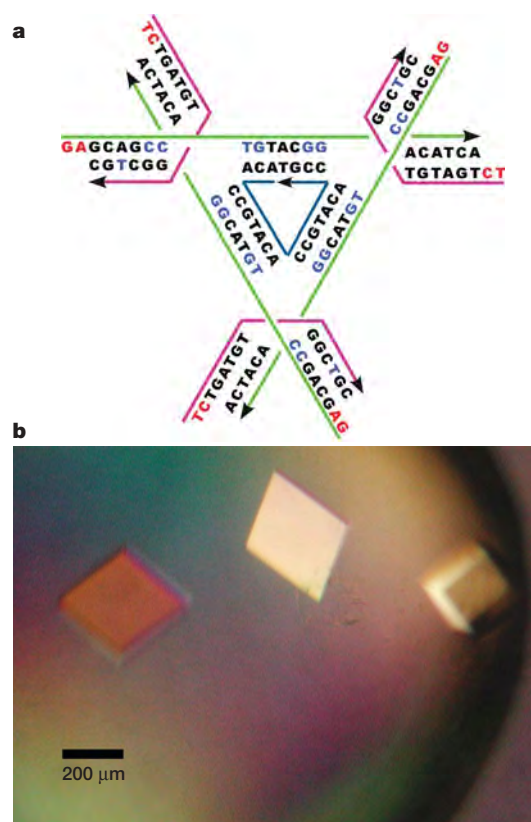


Figure 1 | Schematic design, sequence, and crystal pictures. **a**, Schematic of the tensegrity triangle. The three unique strands are shown in magenta (strands restricted to a single junction), green (strands that extend over each edge of the tensegrity triangle) and dark blue (one unique nicked strand at the centre passing through all three junctions). Arrowheads indicate the 3' ends of strands. Nucleotides with A-DNA-like characteristics are written in bright blue. Cohesive ends are shown in red letters. **b**, An optical image of crystals of the tensegrity triangle. The rhombohedral shape of the crystals and the scale are visible.

¹Department of Chemistry, New York University, New York 10003, USA. ²Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA. ³Structural Biology Center, Argonne National Laboratory, Argonne, Illinois 60439, USA. [†]Present address: Department of Bioengineering, Rice University, 6100 Main Street, MS-142, Houston, Texas 77005, USA.

*These authors contributed equally to this work.

the crystal in a rhombohedral lattice yields a unit cell edge of 69.22 Å and an angle between the edges of 101.44°. Pictures of the rhombohedral crystals are shown in Fig. 1b. Unlike self-assembled two-dimensional crystals, typically a few micrometres in extent⁵, or 3D DNA-nanoparticle crystals^{6,7}, these self-assembled 3D crystals are macroscopic objects, exceeding 250 µm in dimension.

Although the crystallographic asymmetric unit of the unit cell is one-third of the triangle, to comprehend the structure it is most useful to think about the whole triangular structure. The triangle is shown in stereoscopic view in Fig. 2a, where its over-and-under motif is readily visible: For example, the base pairs of the horizontal double helix are in front at the left, but are seen at the rear on the right. Figure 2b shows a view of two of the triangles in electron density, with emphasis on their connection by sticky-ended cohesion. Molecular details are not very reliable at this resolution, but it is clear from the electron density that the DNA is largely in the B-form, with C2'-endo nucleosides. However, seven of the independent nucleotides have characteristics closer to A-form nucleotides rather than B-form, estimated by comparisons with the A–B series of crystal structures in ref. 8; these nucleotides are indicated by blue lettering in Fig. 1a.

The view of the structure perpendicular to a helix axis, seen in Fig. 2b, shows unmistakable double-helical features, such as the major and minor grooves. Holliday junctions, similar to the four-arm junctions that comprise the corners of the triangle, have been examined previously by crystallography⁹ and by atomic force microscopy^{10,11}. The angle between the axes in those cases is about 40–60°, somewhat smaller than the 78° angle enforced by the structure of this tensegrity triangle. A Holliday crossover between two helices is visible at the bottom of Fig. 2a.

This crystal structure demonstrates the viability of designing periodic nucleic acid structures in three dimensions. The three directions that define the lattice are evident from the red, green and yellow colour coding in Fig. 3a, which shows the surroundings of a given tensegrity triangle. The open nature of this stick-like lattice is shown in Fig. 3b, which illustrates the rhombohedron that is flanked by eight of the triangles. The red triangle is at the rear, bonded by sticky ends to the three yellow triangles that flank it lying in a plane closer to the viewer. The yellow triangles are bonded to the green triangles lying in a plane even closer to the viewer. An eighth triangle lying closest to the viewer and directly above the red triangle has been excluded for

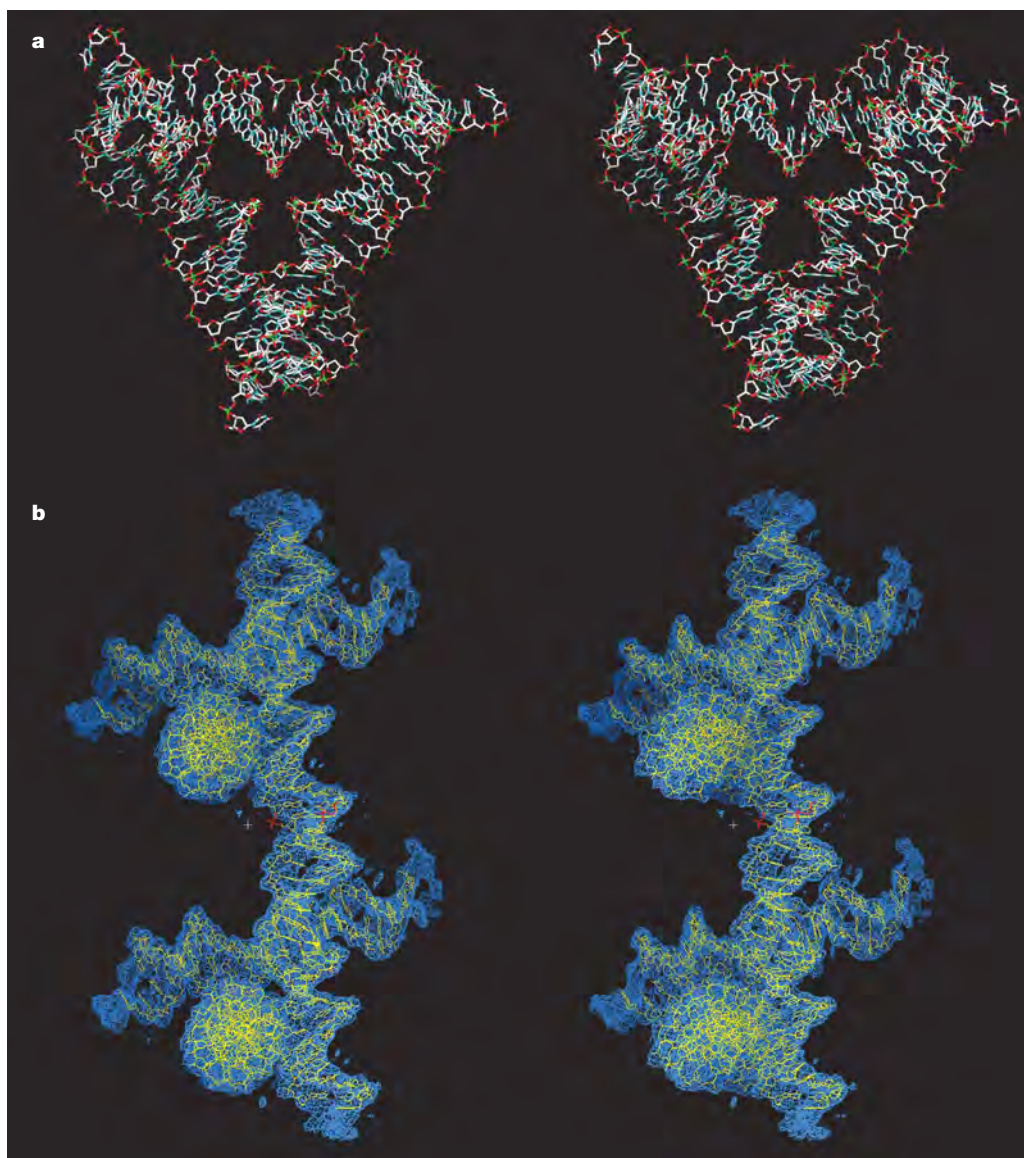


Figure 2 | Views of the tensegrity triangle. **a**, Stereoscopic view of the triangle down its three-fold axis. It is in the same orientation as the schematic in Fig. 1a. The helix on the top edge starts above the mean plane of the molecule at the left and proceeds to the rear as it moves to the right.

b, Stereoscopic view of two triangles in electron density. This image is perpendicular to an edge of the rhombohedron, showing the connection of two triangles by sticky ends. Sticky ends are magenta for emphasis. Some density features belonging to neighbouring molecules are not depicted.

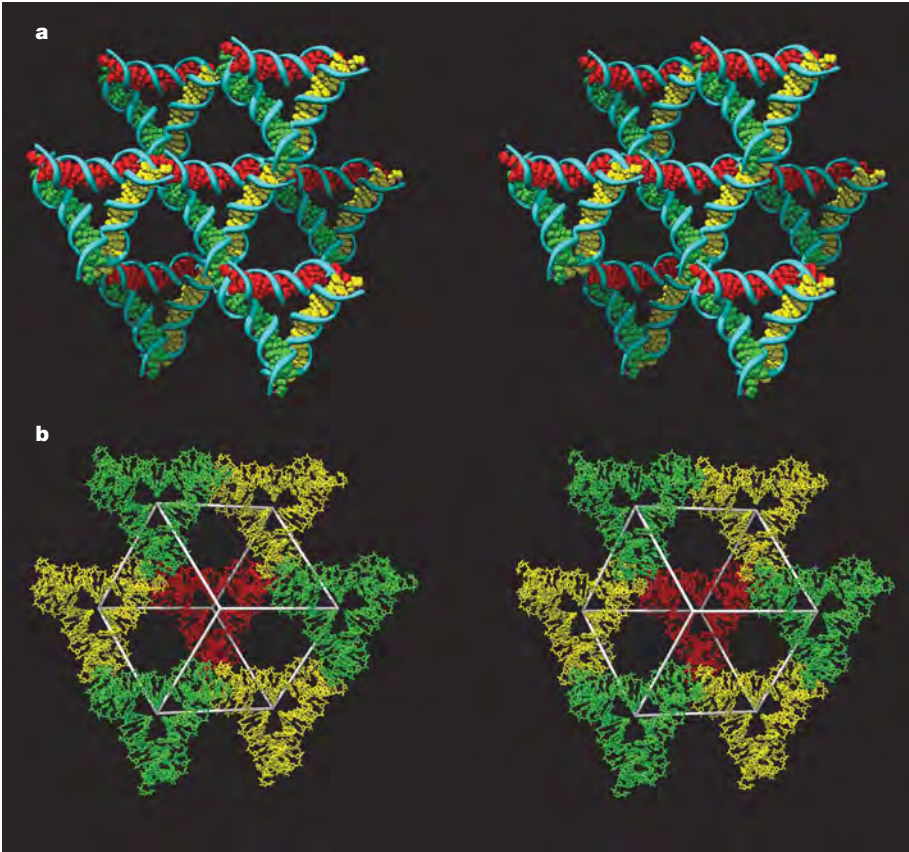


Figure 3 | Lattice formed by tensegrity triangles. **a**, Surroundings of a triangle. This stereoscopic image distinguishes three independent directions by base-pair colour. The central triangle is flanked by six other triangles. **b**, Rhombohedral cavity formed by tensegrity triangles. This stereoscopic image shows seven of the eight triangles that comprise the rhombohedron's

corners. The cavity outline is drawn in white. The rear red triangle connects through one edge each to the three yellow triangles in a plane closer to the viewer. The yellow triangles are connected through two edges each to two different green triangles that are even nearer to the viewer.

clarity. The volume of this rhombohedral cavity is $\sim 103 \text{ nm}^3$, and its open cross-section has an area of $\sim 23 \text{ nm}^2$.

The arrangement of the molecules in this crystal is not fortuitous: it is clearly the result of sticky-ended cohesion as a consequence of our design; indeed, these are the only direct intermolecular contacts in the crystal. As a further demonstration of the ability to program crystalline DNA arrangements in 3D using sticky-ended cohesion, we have constructed eight other rhombohedral lattices from related tensegrity triangles. The crystals are summarized in Table 1, and their structures will be described elsewhere; Table 1 also indicates that these crystals contain cavities that can exceed $1,000 \text{ nm}^3$ (one zepto-litre). The unit cells, lengths and angles are approximately what would be predicted¹² from the design. The resolution of the crystals decreases with increasing edge length; this is possibly a consequence of constructing stick-like lattices that lack the contacts that stabilize

the more common ball-like lattices found in most biomacromolecular crystals. The presence of three-fold rotational symmetry also seems to improve resolution; this is possibly because only three unique strands, rather than seven, need to be purified and mixed with appropriate stoichiometry. Thus, the system described here is a robust basis for crystalline design. A previous system that diffracts well has been reported^{13,14}; however, it does not contain rationally designed Watson–Crick pairing in all directions of cohesion. Designed inorganic¹⁵ 3D periodic systems and coordination networks¹⁶ have been reported, but unlike DNA with sticky ends^{5,17,18}, those systems cannot be used conveniently to design an arbitrary number of components in the asymmetric unit.

The applications that have been suggested previously for designed 3D nucleic acid crystalline systems include the scaffolding of biological systems for crystallographic structure determination¹⁹, as well

Table 1 | Crystalline tensegrity triangle lattices

| Crystal number | Edge length (nucleotide pairs) | Space group | Inter-junction pairs | Rhombohedral cell dimensions | Resolution (Å) | Cross-section (nm ²) | Cavity size (nm ³) |
|----------------|--------------------------------|-------------|----------------------|---|----------------|----------------------------------|--------------------------------|
| 1 | 21 | R3 | 7 | $a = 69.2 \text{ Å}, \alpha = 101.4^\circ$ | 4.0 | 23 | 103 |
| 2 | 21 | P1 | 7 | $a = 68.0 \text{ Å}, \alpha = 102.6^\circ$ | 5.0 | 23 | 101 |
| 3 | 31 | R3 | 17 | $a = 102.0 \text{ Å}, \alpha = 112.7^\circ$ | 6.1 | 62 | 366 |
| 4 | 31 | P1 | 17 | $a = 100.9 \text{ Å}, \alpha = 111.6^\circ$ | 6.3 | 61 | 373 |
| 5 | 32 | R3 | 18 | $a = 103.6 \text{ Å}, \alpha = 113.6^\circ$ | 6.5 | 64 | 367 |
| 6 | 32 | P1 | 18 | $a = 103.3 \text{ Å}, \alpha = 112.2^\circ$ | 6.5 | 64 | 395 |
| 7 | 42 | R3 | 17 | $a = 134.9 \text{ Å}, \alpha = 110.9^\circ$ | 11.0 | 123 | 1,104 |
| 8 | 42 | P1 | 17 | $a = 133.7 \text{ Å}, \alpha = 111.3^\circ$ | 14.0 | 120 | 1,048 |
| 9 | 42 | R3 | 28 | $a = 134.9 \text{ Å}, \alpha = 117.3^\circ$ | 10.0 | 117 | 643 |

The cross-sectional area and cavity size are derived from the lattice parameters. Cross-sections and cavity sizes are estimated by subtracting two radii of the double helix ($\sim 10 \text{ Å}$) from the unit cell dimensions. The space group indicates whether deliberate three-fold rotational averaging has been performed; it has for those in R3, not for those in P1. Edge lengths and inter-junction distances (within triangles) are given in nucleotide pairs. Crystal 1 is the work reported here. The structures of crystals 3 and 7 have been determined by molecular replacement; others are in progress.

as the organization of nanoelectronics²⁰. Both of these applications will probably be most usefully realized with scaffolding that is not three-fold rotationally averaged. Guests that can be accommodated three at a time in the unit cell would be acceptable for structure determination, but they would need to be small enough to fit; three-fold rotational averaging of a single guest would not be optimal for determining structures. Thus, the crystallographic application will probably require higher resolution, and perhaps the larger unit cells noted above, to accommodate larger guest molecules. The nanoelectronics application probably requires only the larger unit cells, so as to accommodate large components. Nevertheless, following this beginning, the other steps apparently needed for these applications are likely to prove incremental and feasible. It seems reasonable that many of the applications of DNA arrays that have been prototyped in two dimensions can be extended to three dimensions. For example, it should be possible to incorporate nanomechanical devices¹⁷ into a 3D lattice, perhaps capable of being programmed to capture specific molecules at specific times^{14,21}. The $\sim 250\text{-}\mu\text{m}$ -sized crystals used in this study contain an estimated 4.5×10^{13} unit cells; metallic nanoparticles have already been used to form patterns on two-dimensional DNA arrays¹⁸, so extending that capability to 3D could provide a very dense ordered arrangement of nanoparticles. In addition to periodic assembly, aperiodic algorithmic assembly²² in three dimensions is also conceivable, but crystallography will not be readily usable for establishing that the assembly has occurred flawlessly.

METHODS SUMMARY

Synthesis, purification and crystallization. DNA sequences were designed using the program SEQUIN²³. DNA strands, including iodinated derivatives, were synthesized by standard phosphoramidite techniques on an Applied Biosystems 394 DNA synthesizer. Strands were doubly purified by reverse-phase high-performance liquid chromatography using a C-18 column (Waters). Crystals were grown from 80- μl sitting drops in a thermally controlled incubator containing $0.25\text{ }\mu\text{g }\mu\text{l}^{-1}$ DNA, 30 mM sodium cacodylate, 50 mM magnesium acetate, 50 mM ammonium sulphate, 5 mM magnesium chloride and 25 mM Tris (pH 8.5), equilibrated against a 1.5-ml reservoir of 1.7 M ammonium sulphate. Rhombohedral-shaped crystals with dimensions as large as $250\text{ }\mu\text{m} \times 250\text{ }\mu\text{m} \times 250\text{ }\mu\text{m}$ were obtained by slow annealing, in which the temperature was decreased from 60 °C to room temperature ($\sim 20\text{ }^{\circ}\text{C}$) with a cooling rate of $0.2\text{ }^{\circ}\text{C}$ per hour over a period of 7 days, during which the volume of the drop diminished by about 90%. Crystals were obtained at the end of the cooling step, and appeared full-sized within a day. Both native and iodinated-derivative crystals were produced under the same conditions.

Data collection. Crystals were transferred to a cryosolvent of 30% glycerol, 100 mM ammonium sulphate, 10 mM MgCl_2 and 50 mM Tris and were frozen by immersion into liquid nitrogen. X-ray diffraction data were collected from crystals of iodinated derivatives (12 iodine atoms per triangle—on the fourth and thirteenth nucleotide of each green strand and on the sixth and eleventh nucleotide of each red strand) at 1.7 Å on beamlines X6A and X25 at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, New York, USA). A complete sphere of native X-ray data was collected at the APS beamline 19ID²⁴.

Structure determination. Complete crystallographic details and associated references are available in the Supplementary Information.

Received 26 February; accepted 6 July 2009.

- Whitesides, G. M., Mathias, J. P. & Seto, C. T. Molecular self-assembly and nanochemistry: A chemical strategy for the synthesis of nanostructures. *Science* **254**, 1312–1319 (1991).
- Seeman, N. C. Nucleic acid junctions and crystal formation. *J. Biomol. Struct. Dyn.* **3**, 11–34 (1985).
- Qiu, H., Dewan, J. C. & Seeman, N. C. A DNA decamer with a sticky end: the crystal structure of d-CGACGATCGT. *J. Mol. Biol.* **267**, 881–898 (1997).
- Liu, D., Wang, W., Deng, Z., Walulu, R. & Mao, C. Tensegrity: construction of rigid DNA triangles with flexible four-arm junctions. *J. Am. Chem. Soc.* **126**, 2324–2325 (2004).
- Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
- Nykypanchuk, D. M. M., Maye, M. M., van der Lelie, D. & Gang, O. DNA-guided crystallization of colloidal nanoparticles. *Nature* **451**, 549–552 (2008).
- Park, S. Y. et al. DNA-programmable nanoparticle crystallization. *Nature* **451**, 553–556 (2008).
- Vargason, J. M., Henderson, K. & Ho, P. S. A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl Acad. Sci. USA* **98**, 7265–7270 (2001).
- Eichmann, B. F., Vargason, J. M., Mooers, B. H. M. & Ho, P. S. The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions. *Proc. Natl Acad. Sci. USA* **97**, 3971–3976 (2000).
- Mao, C., Sun, W. & Seeman, N. C. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J. Am. Chem. Soc.* **121**, 5437–5443 (1999).
- Sha, R., Liu, F. & Seeman, N. C. Atomic force measurement of the inter-domain angle in symmetric Holliday junctions. *Biochemistry* **41**, 5950–5955 (2002).
- Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E. & Seeman, N. C. GIDEON, a program for design in structural DNA nanotechnology. *J. Mol. Graph. Model.* **25**, 470–480 (2006).
- Paukstelis, P. J., Nowakowski, J., Birktoft, J. J. & Seeman, N. C. The crystal structure of a continuous three-dimensional DNA lattice. *Chem. Biol.* **11**, 1119–1126 (2004).
- Paukstelis, P. J. Three dimensional DNA crystals as molecular sieves. *J. Am. Chem. Soc.* **128**, 6794–6795 (2006).
- Furukawa, H., Kim, H. J., Ockwig, N. W., O’Keeffe, M. & Yaghi, O. M. Control of vertex geometry, structure dimensionality, functionality, and pore metrics in the reticular synthesis of crystalline metal-organic frameworks and polyhedra. *J. Am. Chem. Soc.* **130**, 11650–11651 (2008).
- Kawano, M., Kawamichi, T., Haneda, T., Kajima, T. & Fujita, M. The modular synthesis of functional porous coordination networks. *J. Am. Chem. Soc.* **129**, 15418–15419 (2007).
- Ding, B. & Seeman, N. C. Operation of a DNA robot arm inserted into a 2D DNA crystalline substrate. *Science* **314**, 1583–1585 (2006).
- Zheng, J. et al. 2D nanoparticle arrays show the organizational power of robust DNA motifs. *Nano Lett.* **6**, 1502–1504 (2006).
- Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).
- Robinson, B. H. & Seeman, N. C. The design of a biochip: a self-assembling molecular-scale memory device. *Protein Eng.* **1**, 295–300 (1987).
- Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. Dynamic patterns programmed by DNA tiles captured on a DNA origami substrate. *Nature Nanotech.* **4**, 245–249 (2009).
- Rothmund, P. W. K., Papadakis, N. & Winfree, E. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol.* **2**, 2041–2053 (2004).
- Seeman, N. C. *De novo* design of sequences for nucleic acid structure engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581 (1990).
- Rosenbaum, G. et al. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J. Synchrotron Radiat.* **13**, 30–45 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research has been supported by grants to N.C.S. from the National Institute of General Medical Sciences, the National Science Foundation, the Army Research Office, the Office of Naval Research and the W. M. Keck Foundation. It has also been supported by NSF grant CCF-0622093 and NIH grant 1R21EB007472 to C.M. We thank W. Sherman for assistance in establishing the likely structural features of tensegrity triangles. We thank R. Sweet, M. Allaire, H. Robinson, A. Saxena and A. Héroux at the BNL-NSLS at beamlines X6A and X25 of the National Synchrotron Light Source. BNL-NSLS is supported principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. The use of the 19ID beamline at the Structural Biology Center/Advanced Photon Source is supported by the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Author Contributions J.Z. grew crystals, collected data, analysed data and wrote the paper; J.J.B. collected data, analysed data and wrote the paper; Y.C. grew crystals, collected data and analysed data; T.W. grew crystals, collected data, analysed data and wrote the paper; R.S. grew crystals, analysed data and wrote the paper; P.E.C. grew crystals and analysed data; S.L.G. collected data and analysed data; C.M. devised the motif, analysed data and wrote the paper; N.C.S. initiated the project, analysed data and wrote the paper.

Author Information Atomic coordinates and experimental structure factors have been deposited within the Protein Data Bank and are accessible under the code 3GBI. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.C.S. (ned.seeman@nyu.edu) or C.M. (mao@purdue.edu).

LETTERS

Denitrification as the dominant nitrogen loss process in the Arabian Sea

B. B. Ward¹, A. H. Devol², J. J. Rich³, B. X. Chang², S. E. Bulow¹, Hema Naik⁴, Anil Pratihary⁴ & A. Jayakumar¹

Primary production in over half of the world's oceans is limited by fixed nitrogen availability. The main loss term from the fixed nitrogen inventory is the production of dinitrogen gas (N_2) by heterotrophic denitrification or the more recently discovered autotrophic process, anaerobic ammonia oxidation (anammox). Oceanic oxygen minimum zones (OMZ) are responsible for about 35% of oceanic N_2 production and up to half of that occurs in the Arabian Sea¹. Although denitrification was long thought to be the only loss term, it has recently been argued that anammox alone is responsible for fixed nitrogen loss in the OMZs^{2–4}. Here we measure denitrification and anammox rates and quantify the abundance of denitrifying and anammox bacteria in the OMZ regions of the Eastern Tropical South Pacific and the Arabian Sea. We find that denitrification rather than anammox dominates the N_2 loss term in the Arabian Sea, the largest and most intense OMZ in the world ocean. In seven of eight experiments in the Arabian Sea denitrification is responsible for 87–99% of the total N_2 production. The dominance of denitrification is reproducible using two independent isotope incubation methods. In contrast, anammox is dominant in the Eastern Tropical South Pacific OMZ, as detected using one of the isotope incubation methods, as previously reported^{3,5}. The abundance of denitrifying bacteria always exceeded that of anammox bacteria by up to 7- and 19-fold in the Eastern Tropical South Pacific and Arabian Sea, respectively. Geographic and temporal variability in carbon supply may be responsible for the different contributions of denitrification and anammox in these two OMZs. The large contribution of denitrification to N_2 loss in the Arabian Sea indicates the global significance of denitrification to the oceanic nitrogen budget.

Nitrogen limits biological production in many terrestrial and aquatic environments, so it is essential to understand processes that control nitrogen availability. The global inventory of fixed nitrogen may be out of balance, with losses exceeding inputs⁶. This perceived imbalance has motivated research into inputs via nitrogen fixation and losses via denitrification and anammox. The discoveries of previously unknown marine microbes involved in marine nitrogen fixation⁷ and that anammox may be the dominant N_2 loss term in many environments^{5,8,9} both point to large uncertainties in our understanding of the global ocean nitrogen cycle and thus our ability to predict how it might respond to perturbation.

Denitrification is an anaerobic respiratory process that is found widely in both autotrophic and heterotrophic microbes in all three biological domains. Phylogenetic studies of functional genes involved in denitrification indicate that both the Eastern Tropical South Pacific (ETSP) and Arabian Sea OMZs (defined as the depth zone where oxygen concentrations are low enough to induce anaerobic metabolism) harbour large diverse assemblages of heterotrophic denitrifying bacteria^{10–12}. In the absence of oxygen, denitrifiers respire nitrate

(NO_3^-) sequentially to nitrite (NO_2^-), nitric and nitrous oxides (NO , N_2O) and finally to N_2 . In consuming organic matter, they regenerate inorganic nutrients such as carbon dioxide (CO_2), ammonium (NH_4^+), and phosphate (PO_4^{3-}), which are necessary to sustain continued primary production. Anammox bacterial assemblages in nature are less diverse than denitrifier assemblages and are represented by only one or two phylotypes^{12–14}. In contrast to denitrifying bacteria, anammox bacteria are autotrophs that consume NH_4^+ and NO_2^- in respiration and form biomass by CO_2 fixation. Anammox is thus dependent upon nutrient regeneration by some other process in order to supply the inorganic nutrients and reductants it requires. Denitrification by heterotrophic bacteria seemed to be the obvious supply process for both NH_4^+ and NO_2^- , which are required by anammox, and unlike the wastewater treatment plants in which anammox was originally identified, are in very low concentrations in most of the ocean. It is therefore difficult to understand how anammox could occur in the ETSP OMZ to the exclusion of denitrification^{3,5}.

Lam *et al.*⁴ suggested that dissimilatory nitrate reduction to ammonium is a potential source of the NH_4^+ in the ETSP and argued that other processes can supply most of the nitrite. A revision of the nitrogen cycle of OMZs was proposed, in which conventional denitrification is unnecessary because nitrite is supplied by either aerobic nitrifying or anaerobic nitrate-reducing microbes and ammonium is supplied by dissimilatory nitrate reduction to ammonium. This revision is difficult to reconcile with numerous reports of abundant and diverse denitrifying bacterial assemblages in OMZs, where their facultative anaerobic metabolism should be favoured by plentiful organic supply and lack of oxygen. Although not detected in short-term incubations in the ETSP^{3,5}, denitrification has, however, been detected in longer incubations in small (12 ml) glass vials called exetainers elsewhere¹⁵ and in the ETSP and Arabian Sea in much larger volume containers^{16,17}. These observations provide evidence that viable denitrifying bacteria are present in the major OMZs and this, along with the need for nitrogen substrates to support anammox, seems incompatible with the lack of measurable denitrification. To reconcile these findings, we measured denitrification and anammox rates and quantified the abundance of denitrifying and anammox bacteria in the OMZ regions of the ETSP and the Arabian Sea (Supplementary Fig. 1).

All four stations in both systems exhibited the classic oxygen and dissolved inorganic nitrogen profiles usually associated with OMZs (Fig. 1). In the Arabian Sea pelagic OMZ, we measured denitrification and anammox using two methods: (1) the exetainer method previously used in the ETSP¹⁵ (8 ml volume, 36–48 h incubations, helium-sparged) and (2) incubations in large trace-metal clean tri-laminate gas-impermeable bags¹⁸ (~8-litre volume with no headspace, 48 h incubations, ambient gas concentrations). No changes in dissolved inorganic nitrogen concentrations were detectable by

¹Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA. ²Department of Oceanography, University of Washington, Seattle, Washington 98195, USA. ³Center for Environmental Studies, Brown University, Providence, Rhode Island 02912, USA. ⁴National Institute of Oceanography, Dona Paula, Goa 403004, India.

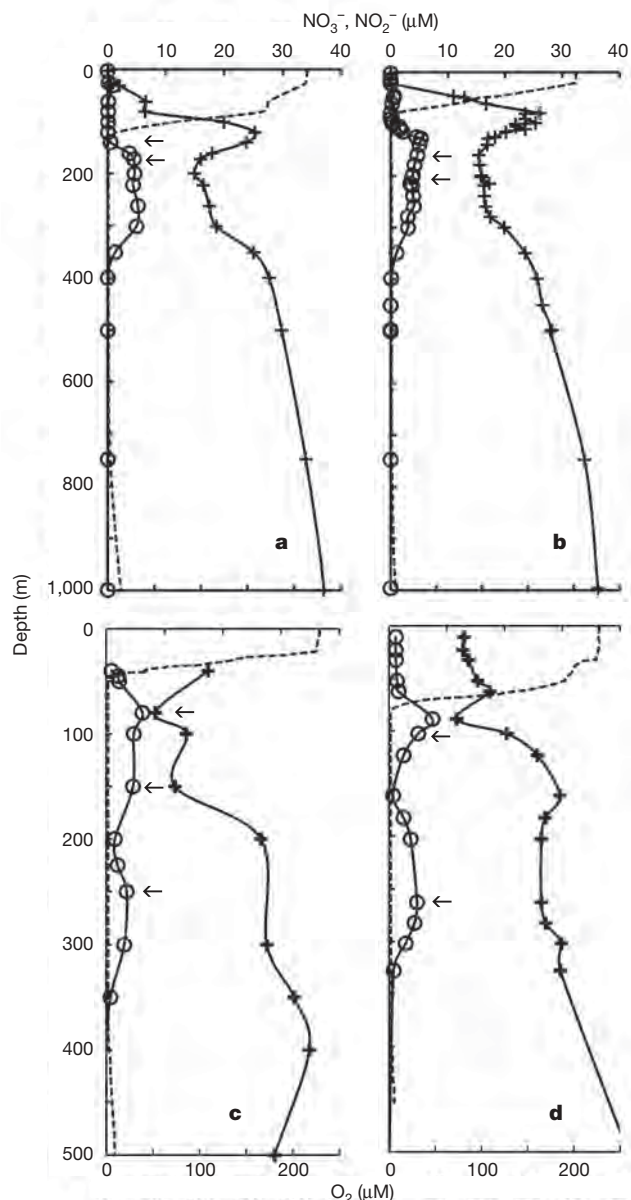


Figure 1 | Chemical profiles for experimental stations in the Arabian Sea and ETSP. Oxygen (dotted traces), nitrate (crosses), and nitrite (circle) concentrations at two stations in the Arabian Sea (**a, b**) and two in the ETSP off Peru (**c, d**). Solid lines are smoothed curves connecting the data. **a**, Station 1, 19° N, 67° E, bottom depth 3,100 m. **b**, Station 2, 15° N, 64° E, bottom depth 3,930 m. **c**, Station 9, 15° 38' S, 75° 08' W, bottom depth 977 m. **d**, Station 24, 12° 15' S, 79° 18' W, bottom depth 4,899 m. Arrows indicate depths where denitrification and anammox rates were measured.

spectrophotometric methods over the course of the incubations, and transformation rates were detectable only by isotopic measurements of the N_2 pool ($^{14,15}N_2$ (that is, $^{14}N \equiv ^{15}N$) production from $^{15}NH_4^+$ as a tracer for anammox and $^{15}N_2$ production from $^{15}NO_2^-$ for denitrification). Denitrification rates ranged from 0.24 to 25.4 nM of N_2 per day in the exetainers and from 1.3 to 6.5 nM of N_2 per day in the bags (Fig. 2). Anammox rates were generally much lower, 0.24–4.32 nM N_2 per day in the exetainers and 0.12–0.21 nM N_2 per day in the bags. Although the bag and exetainer incubations were performed on different days with water collected from different casts and incubated under different conditions, the difference in rates obtained by the two methods is relatively minor. Denitrification rates calculated from $^{15}N_2$ production include the coupled pathway of dissimilatory

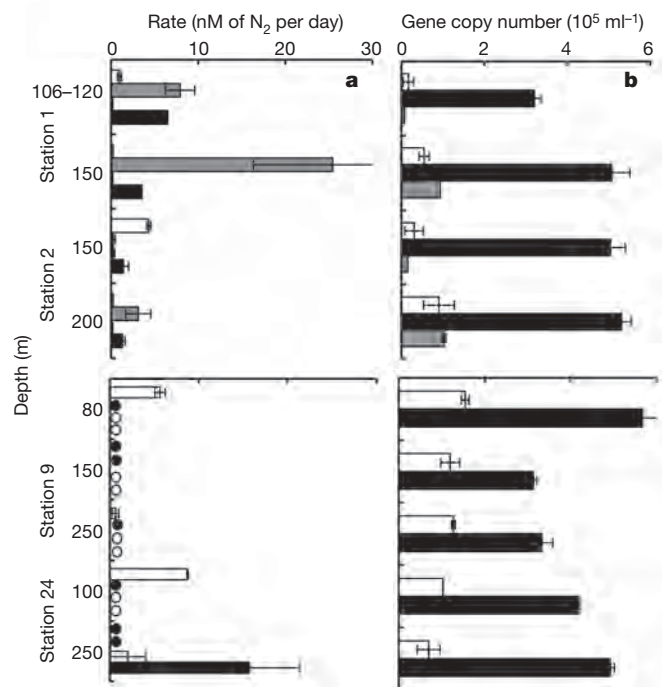


Figure 2 | Rates of denitrification and anammox and abundance of signature genes for denitrification and anammox. Top, Arabian Sea; bottom, ETSP. **a**, Denitrification and anammox rates. White bars, anammox rate measured in exetainers; grey bars, denitrification rate measured in exetainers; hatched bars, anammox rate measured in 8-litre bags; black bars, denitrification rate measured in 8-litre bags; black circle, no rate detected; white circle, not measured. Error bars are the standard error of the slope of $^{14,15}N_2$ or $^{14,15}N_2$ plus $^{15}N_2$ accumulation in exetainer incubations. Error bars indicate error of the slope for exetainer incubations and range of duplicates for bag incubations. **b**, White bars, depth distribution of anammox 16S rRNA genes; black bars, total denitrifier *nirS* genes; grey bars, a dominant *nirS* clade. Error bars indicate standard deviations of triplicate Q-PCR assays.

nitrate reduction to ammonium and anammox, but the rate of this coupled pathway cannot exceed the anammox rate. For the seven experiments excluding the high anammox rate from 150 m at station 2, denitrification averaged 93.7% of the total N_2 production rate.

To quantify the abundance of signature genes representing denitrification, we used degenerate primers¹⁹ for the *nirS* gene (which encodes the haem-containing nitrite reductase gene in the denitrification pathway), and for anammox, we used anammox-specific 16S ribosomal RNA gene primers²⁰. *nirS* abundance measured by quantitative PCR (Q-PCR) in the Arabian Sea was $(3.2\text{--}5.3) \times 10^5$ copies ml^{-1} , and exceeded anammox abundance by 5.9–19.2-fold at the two stations (stations 1 and 2) for which anammox rates are reported (Fig. 2). In 2004, near the same station locations, we investigated the diversity of the *nirS* genes and found that one clade was dominant at station 1. This clade, Dom *nirS*, represented 25% of the sequences in clone libraries and 30% of the total *nirS* abundance by Q-PCR near station 1 (ref. 12). In 2007, at the time of our rate measurements, the Dom *nirS* clade was present at up to 20% of the total *nirS* by Q-PCR, similar to the abundance of the anammox 16S rRNA genes. Neither the total *nirS* primers¹⁹ nor the Dom *nirS* primers¹² amplify the known anammox *nirS* sequences, so these abundances do not include anammox genes. Assuming both *nirS* and the anammox 16S rRNA genes are present in one copy per cell^{21,22}, *nirS*-type denitrifiers represent a large fraction of the total cells present^{12,23} and are more numerous than anammox bacteria.

In the ETSP, only anammox was detected using the exetainer incubation method (J.J.R., B.X.C., A.H.D. & B.B.W., manuscript in

preparation), although anammox was not detected at all depths (Fig. 2). No $^{15}\text{N}_2$ (indicating denitrification) was detected in incubations with $^{15}\text{NO}_3^-$ added as a tracer. The range in anammox rates and the lack of denitrification were consistent with previous reports for the ETSP^{3,5} in which the same methods were used. Denitrification, at a rate similar to rates measured in the bag incubations in the Arabian Sea, was measured in bag incubations at 260 m at station 24 in the ETSP, although exetainer experiments detected neither anammox nor denitrification at this depth (Fig. 2).

At the same depths in the ETSP where anammox or denitrification was detected, the abundance of denitrifiers containing the *nirS* gene was $(3.1\text{--}5.7) \times 10^5$ cells ml^{-1} . The Dom *nirS* clade from the Arabian Sea was not detected by Q-PCR in any samples from the ETSP. Published clone library data^{4,10} so far have not identified a clear dominant phylotype in the ETSP system. The abundance of anammox bacteria in the ETSP samples was $(0.7\text{--}1.5) \times 10^5$ cells ml^{-1} , $\sim 3\text{--}7$ -fold less abundant than the *nirS* denitrifiers. At stations 3° north of our station 9, Hamersley *et al.*⁵ reported abundances of anammox bacteria based on Q-PCR and fluorescent in-situ hybridization (FISH) (targeting the 16S rRNA sequence) of up to 1.5×10^5 cells ml^{-1} . The similarity between the anammox abundances reported here (Fig. 2) and those reported previously for the same region⁵ lends support to the general pattern we observed. The similarity in *nirS* denitrifier abundances between the Arabian Sea and ETSP suggests that denitrifiers are important in both systems, but anammox abundance was generally less in the Arabian Sea, where anammox rates were much lower.

Denitrifier abundances reported here are probably underestimates of the total denitrifier abundance because the *nirS* primers may not detect all *nirS*-containing organisms and denitrifiers that use the copper-type nitrite reductase, encoded by the *nirK* gene, were not quantified. The *nirS* primers used by Lam *et al.*⁴ significantly underestimate the denitrifier abundance, because they do not amplify most of the sequenced *nirS* genes reported previously from OMZ waters in the Arabian Sea¹² and the ETSP¹⁰.

The rates reported here are the first direct measurements of denitrification in the OMZ of the Arabian Sea in short-term incubations. Denitrification, defined as production of $^{14,15}\text{N}_2$ from $^{15}\text{NO}_3^-$ in large-volume incubations reported previously¹, could not distinguish between anammox and denitrification. Nicholls *et al.*¹⁷ inferred denitrification in the Arabian Sea from rates of production of N_2O from $^{15}\text{NO}_2^-$ and $^{15}\text{NO}_3^-$ (and rates of production of $^{15}\text{N}_2$ from $^{15}\text{NO}_2^-$, mainly after at least two days); this group also detected only very low rates of anammox. In our study, both exetainers and bags yielded rates of similar magnitude in incubations of similar length (up to two days). The range of denitrification rates reported here includes the average rate of 9.1 ± 1.1 nM N_2 per day reported for total denitrification (implicitly including anammox)¹ in the Arabian Sea, the mean rate of 2.0 nM N_2 per day for $^{15}\text{N}_2$ production from $^{15}\text{NO}_2^-$ in longer incubations in the Arabian Sea¹⁷, and the rate of 14.7 ± 3.53 nM N_2 per day estimated from N_2O consumption in the ETSP²⁴.

Denitrification rates were much faster than anammox at all depths and all stations except in one exetainer experiment. The significant rates of denitrification that we measured in the Arabian Sea contrast with rates measured in the ETSP by ourselves and others, in which anammox, but not denitrification, was detected^{3,5}. Because identical methods were used in the ETSP and Arabian Sea, we suggest that the Arabian Sea differs from the ETSP in the relative contribution of denitrification and anammox, at least at the times of these experiments. We previously documented that denitrification was limited by availability of organic carbon in the OMZ of the ETSP, but not in the Arabian Sea¹⁶. Therefore it seems likely that important differences in carbon availability, in terms of magnitude, composition or temporal supply, may be responsible for these ecosystem-scale differences²⁵ and might also lead to temporally and spatially dynamic distributions of both processes.

The high abundance and diversity of denitrifiers, compared to the lower abundance and lower diversity of anammox organisms, suggests that dynamic assemblages of denitrifiers persist in both OMZs. It is clear, however, that the denitrifier assemblages in the two environments differ in important ways, including the consistent detection of a dominant denitrifier phylotype, which was absent in the ETSP, but which was associated with the core of the OMZ and highest denitrification rates in the Arabian Sea.

The lack of correlation between bacterial abundances and the respective rates is compatible with the contrasting lifestyles of the two kinds of microbes: versatile opportunistic heterotrophic denitrifying bacteria are probably capable of a much more dynamic activity response to episodic substrate supply than are the more highly constrained slower-growing autotrophic anammox bacteria. We hypothesize that episodic organic matter supply supports the denitrifiers in the ETSP and that at times of high organic flux, the relative contribution of denitrification and anammox shifts towards the ratio observed in the Arabian Sea. The close coupling between aerobic nitrification and anammox, which might be able to supply some of the nitrite required for anammox in the ETSP, is unlikely to be important in the core depths of the OMZ in either the ETSP or the Arabian Sea. When detected, the highest anammox rates generally occur near the upper edge of the OMZ in the ETSP⁵, whereas we detected denitrification in the core of the OMZ at the depth of the secondary nitrite maximum in the Arabian Sea and in the ETSP. The conventional nitrogen cycle of the Arabian Sea must be revised to include anammox, but denitrification is clearly the overwhelmingly dominant flux for fixed nitrogen loss in the ocean's largest OMZ. Its importance in the Arabian Sea, which alone is responsible for up to half of the fixed nitrogen loss from the pelagic realm¹, indicates that denitrification is a major process in the overall marine nitrogen cycle.

METHODS SUMMARY

Sampling was conducted aboard the RV *Knorr* in the ETSP in 2005 and the RV *Roger Revelle* in the Arabian Sea in 2007. Nutrient concentrations were determined by standard spectrophotometric methods and oxygen was measured using an electrode mounted on the conductivity–temperature–depth instrument package. Water for incubations was collected from Niskin bottles, which were plumbed with helium during sample removal to avoid atmospheric contamination. Anammox and denitrification rates were measured using the standard method with small-volume exetainers¹⁵, in which various isotope additions were made to helium-sparged vials and the accumulation of $^{15}\text{N}_2$ was measured on a Europa 20/20 mass spectrometer in triplicate vials poisoned at 12, 24, 36, and 48 h. Exactly the same exetainer methods were used as described in ref. 15 to obtain results that were exactly comparable to those previously reported. Anammox was quantified from the accumulation of $^{14,15}\text{N}_2$ in the incubation to which $^{15}\text{NH}_4^+$ was added as a tracer along with $^{14}\text{NO}_2^-$, and denitrification was quantified as the accumulation of $^{14,15}\text{N}_2$ and $^{15}\text{N}_2$ in incubations to which $^{15}\text{NO}_2^-$ was added as a tracer. Rates of denitrification and anammox were considered significant if the increase in $^{14,15}\text{N}_2$ or $^{15}\text{N}_2$ was linear over the first 36 h of exetainer incubations. The new method used large trilaminate bags with no headspace and no sparging¹⁶. The same tracer treatments as in the exetainer method were used but at lower concentrations. Degassed tracer solutions were added while filling the bags directly from the Niskin sampling bottles. Single endpoint determinations of $^{14,15}\text{N}_2$ and $^{15}\text{N}_2$ were made on subsamples from the 8-litre bags after 48 h and analysed following the method of Emerson *et al.*²⁶ on a Finnigan Delta XL mass spectrometer. Signature genes for denitrification and anammox were quantified by Q-PCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 8 July 2009.

1. Devol, A. H. *et al.* Denitrification rates and excess nitrogen gas concentrations in the Arabian Sea oxygen deficient zone. *Deep-Sea Res.* **53**, 1533–1547 (2006).
2. Kuypers, M. M. M. *et al.* Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc. Natl Acad. Sci. USA* **102**, 6478–6483 (2005).
3. Thamdrup, B. *et al.* Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnol. Oceanogr.* **51**, 2145–2156 (2006).

4. Lam, P. *et al.* Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Natl Acad. Sci. USA* **106**, 4752–4757 (2009).
5. Hamersley, M. R. *et al.* Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. *Limnol. Oceanogr.* **52**, 923–933 (2007).
6. Codispoti, L. *et al.* The oceanic fixed nitrogen and nitrous oxide budgets: moving targets as we enter the Anthropocene? *Sci. Mar.* **65**, 85–105 (2001).
7. Zehr, J. P. *et al.* Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* **412**, 635–638 (2001).
8. Kuypers, M. M. *et al.* Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc. Natl Acad. Sci. USA* **102**, doi:10.11073/pnas.0502088102 (2005).
9. Dalsgaard, T., Thamdrup, B. & Canfield, D. E. Anaerobic ammonium oxidation (anammox) in the marine environment. *Res. Microbiol.* **156**, 457–464 (2005).
10. Castro-Gonzalez, M., Braker, G., Farias, L. & Ulloa, O. Communities of *nirS*-type denitrifiers in the water column of the oxygen minimum zone in the eastern South Pacific. *Environ. Microbiol.* doi:10.1111/j.1462-2920 (2005).
11. Jayakumar, D. A., Francis, C. A., Naqvi, S. W. A. & Ward, B. B. Diversity of nitrite reductase genes (*nirS*) in the denitrifying water column of the coastal Arabian Sea. *Aq. Microb. Ecol.* **34**, 69–78 (2004).
12. Jayakumar, A., Naqvi, S. W. A. & Ward, B. B. in *Indian Ocean Biogeochemical Processes and Ecological Variability* (eds Wiggert, J. D., Hood, R. R., Naqvi, S. W. A., Brink, K. H., & Smith, S. L.) (AGV Geophysical Monograph series, American Geophysical Union, in the press).
13. Schmid, M. C. *et al.* Anaerobic ammonium-oxidizing bacteria in marine environments: widespread occurrence but low diversity. *Environ. Microbiol.* **9**, 1476–1484 (2007).
14. Woebken, D. *et al.* A microdiversity study of anammox bacteria reveals a novel *Candidatus Scalindua* phylotype in marine oxygen minimum zones. *Environ. Microbiol.* **10**, 3106–3119 (2008).
15. Dalsgaard, T., Canfield, D. E., Petersen, J., Thamdrup, B. & Acuna-Gonzalez, J. N_2 production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica. *Nature* **422**, 606–608 (2003).
16. Ward, B. B. *et al.* Organic carbon, and not copper, controls denitrification in oxygen minimum zones of the ocean. *Deep-Sea Res. I.* doi:10.1016/j.dsr.2008.07.005 (2008).
17. Nicholls, J. C., Davies, I. M. & Trimmer, M. High-resolution profiles and nitrogen isotope tracing reveal a dominant source of nitrous oxide and multiple pathways of nitrogen gas formation in the central Arabian Sea. *Limnol. Oceanogr.* **52**, 156–168 (2007).
18. Ward, B. B. *et al.* Denitrification in the hypolimnion of permanently ice-covered Lake Bonney, Antarctica. *Aq. Microb. Ecol.* **38**, 295–307 (2005).
19. Braker, G., Fesefeldt, A. & Witzel, K.-P. Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. *Appl. Environ. Microbiol.* **64**, 3769–3775 (1998).
20. Schmid, M. *et al.* *Candidatus "Scalindua broda"*, sp. nov., *Candidatus "Scalindua wagneri"*, sp. nov., two new species of anaerobic ammonia oxidizing bacteria. *Syst. Appl. Microbiol.* **26**, 529–538 (2003).
21. Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
22. Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
23. Jayakumar, A., O'Mullan, G. D., Naqvi, S. W. A. & Ward, B. B. Denitrifying bacterial community composition changes associated with stages of denitrification in oxygen minimum zones. *Microb. Ecol.* **58**, 350–362 (2009).
24. Farias, L. *et al.* Denitrification and nitrous oxide cycling within the upper oxycline of the eastern tropical South Pacific oxygen minimum zone. *Limnol. Oceanogr.* **54**, 132–144 (2009).
25. Ducklow, H. W. *et al.* Heterotrophic bacterioplankton in the Arabian Sea: Basinwide response to year-round high primary productivity. *Deep-Sea Res. II* **48**, 1303–1323 (2001).
26. Emerson, S. E., Stump, C., Wilbur, D. & Quay, P. Accurate measurement of O_2 , N_2 , and Ar gases in water and the solubility of N_2 . *Mar. Chem.* **74**, 337–347 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements W. Naqvi helped make it possible to schedule the RV *Roger Revelle* to work in the Arabian Sea in 2007. We thank the master and crews of the RV *Knorr* and RV *Roger Revelle*, and J. Moffett who was chief scientist on the *Knorr*. This work was supported by NSF grants to B.B.W., A.H.D. and A.J.

Author Contributions B.B.W., A.H.D. and A.J. designed the Arabian Sea experiments and participated in all the fieldwork; J.J.R. designed the ETSP and Arabian Sea exetainer experiments and performed them in the ETSP; S.E.B. and B.B.W. performed the exetainer experiments in the Arabian Sea; J.J.R. made the exetainer mass spec measurements for both Arabian Sea and ETSP; B.X.C. and A.H.D. made the mass spectrometer measurements for the bag experiments; B.B.W., A.H.D., A.J., S.E.B. and B.X.C. carried out the bag experiments in the Arabian Sea; A.J. performed the Q-PCR experiments; H.N. and A.P. provided the nitrite and nitrate data for Fig. 1a and b; B.B.W. wrote the paper with input from A.H.D., A.J., S.E.B., B.X.C. and J.J.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.B.W. (bbw@princeton.edu).

METHODS

Inorganic nutrient concentrations were measured by autoanalyser and CTD oxygen values were standardized to Winkler titrations. Methods of DNA extraction, PCR amplification and Q-PCR using SYBR Green for *nirS* and *Dom nirS*^{11,12} and anammox 16S rRNA genes²⁰ (primers AMX368F and AMX820R) have been described previously. Standardization and verification of specificity for Q-PCR assays were performed as previously described¹².

Denitrification and anammox rates in small volumes were measured using the previously described incubation protocol¹⁵. Water samples were collected from Niskin bottles, which were plumbed with helium during sample removal to avoid atmospheric contamination of gases and held at 12 °C (<3 h), then transferred to a N₂ flushed glove bag, where the ¹⁵N tracers (3 µM final concentration ¹⁵NO₂[−] or ¹⁵NH₄⁺) were added. Triplicate aliquots (8 ml) were incubated in 12-ml exetainers, which were sealed and sparged with helium for 5 min. Three replicates of each sample were killed with 0.05 ml of 7 M ZnCl₂ at 12-h intervals between 0 and 48 h.

The amount of excess ¹⁵N₂ produced during the incubations was measured directly in the incubation vials with a continuous-flow isotope ratio mass spectrometer (Europa Scientific 20-20), in-line with an automated gas preparation unit (Europa Scientific, ANCA-G Plus). Denitrification was defined as binomially distributed ^{14,15}N₂ and ¹⁵N₂ production in the ¹⁵NO₂[−] incubation²⁷ and anammox was defined as ^{14,15}N₂ production in ¹⁵NH₄⁺ incubation²⁸. Statistical analyses were performed using the JMP software program (SAS). The ¹⁵N₂ production rates were calculated from the 36-h (four time points) slope of the standard least-squares regression. Error was calculated as the standard error of the slope. Error for denitrification rates was determined by propagation of error from the slopes of F29 and F30 (according to ref. 29).

Isotope measurements for bag incubations were made on subsamples collected from the bags into HgCl₂-poisoned, pre-evacuated, 300-ml glass flasks equipped with gas-tight 9-mm-bore Louwers-Hapert single o-ring valves. The amount of ¹⁴N₂, ^{14,15}N₂ or ¹⁵N₂ added by denitrification in the ¹⁵NO₂[−]-amended incubations was calculated using the following equations:

$$[^{15}\text{N}_2]_{\text{denit}} = \frac{[^{15}\text{N}_2]_{\text{initial}} - R_{\text{final}}^{15} \times [^{14}\text{N}_2]_{\text{initial}}}{R_{\text{final}}^{15} \times \left(\frac{P_{\text{denit}}^{14}}{P_{\text{denit}}^{15}} \right) - 1}$$

$$[^{14}\text{N}_2]_{\text{denit}} = [^{15}\text{N}_2]_{\text{denit}} \times \left(\frac{P_{\text{denit}}^{14}}{P_{\text{denit}}^{15}} \right)$$

$$[^{14,15}\text{N}_2]_{\text{denit}} = [^{15}\text{N}_2]_{\text{denit}} \times \left(\frac{P_{\text{denit}}^{14,15}}{P_{\text{denit}}^{15}} \right)$$

The denitrification rate is $[^{14}\text{N}_2]_{\text{denit}} + [^{14,15}\text{N}_2]_{\text{denit}} + [^{15}\text{N}_2]_{\text{denit}}$, divided by time (in days), where $[^x\text{N}_2]_{\text{denit}}$ is the amount of ^xN₂ added by denitrification, $[^x\text{N}_2]_{\text{initial}}$ is the initial amount of ^xN₂, $R_{\text{final}}^x = ^x\text{N}_{2,\text{final}} / ^{14}\text{N}_{2,\text{final}}$ is the isotope ratio at the endpoint, and P_{denit}^x is the probability of making ^xN₂ by denitrification.

The amount of ¹⁴N₂ or ^{14,15}N₂ added by anammox in the ¹⁵NH₄⁺-amended incubations, in which there was no measurable ambient ¹⁴NH₄⁺, was calculated using the following equations:

$$[^{14,15}\text{N}_2]_{\text{ana}} = [^{14}\text{N}_2]_{\text{initial}} \times R_{\text{final}}^{14,15} - [^{14,15}\text{N}_2]_{\text{initial}}$$

$$[^{14}\text{N}_2]_{\text{ana}} = [^{14,15}\text{N}_2]_{\text{ana}} \times \left(\frac{P_{\text{ana}}^{14}}{P_{\text{ana}}^{14,15}} \right)$$

The anammox rate is $[^{14}\text{N}_2]_{\text{ana}} + [^{14,15}\text{N}_2]_{\text{ana}}$, divided by time (in days), where $[^x\text{N}_2]_{\text{ana}}$ is the amount of ^xN₂ added by anammox, and P_{ana}^x is the probability of making ^xN₂ by anammox.

27. Nielsen, L. Denitrification in sediment determined from nitrogen isotope pairing. *FEMS Microbiol. Ecol.* **86**, 357–362 (1992).

28. Thamdrup, B. & Dalsgaard, T. Production of N₂ through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Appl. Environ. Microbiol.* **68**, 1312–1318 (2002).

LETTERS

The oldest hand-axes in Europe

Gary R. Scott¹ & Luis Gibert¹

Stone tools are durable reminders of the activities, skills and customs of early humans, and have distinctive morphologies that reflect the development of technological skills during the Pleistocene epoch. In Africa, large cutting tools (hand-axes and bifacial chopping tools) became part of Palaeolithic technology during the Early Pleistocene (~1.5 Myr ago)^{1–3}. However, in Europe this change had not been documented until the Middle Pleistocene (<0.5 Myr ago)^{4,5}. Here we report dates for two western Mediterranean hand-axe sites that are nearly twice the age of the supposed earliest Acheulian in western Europe. Palaeomagnetic analysis of these two sites in southeastern Spain found reverse polarity magnetozones, showing that hand-axes were already in Europe as early as 0.9 Myr ago. This expanded antiquity for European hand-axe culture supports a wide geographic distribution of Palaeolithic bifacial technology outside of Africa during the Early Pleistocene.

Africa has an extensive Pliocene–Pleistocene record of hominins, with the oldest archaeological sites (~1.5 Myr old) that contain large bifacial cutting tools (Acheulian hand-axes and cleavers) found at Konso-Gardula¹, Peninj² and Wonderwerk³. There are few early archaeological sites outside Africa with Acheulian bifacial artefacts, mostly around the Early/Middle Pleistocene boundary, such as Gesher Benot Ya'aqov, Israel (palaeomagnetic age: 0.7 Myr old)⁶, and Bose, China (⁴⁰Ar–³⁹Ar age: 0.8 Myr old)⁷. An exception is the older, eastern Mediterranean site of 'Ubeidiya, Israel (palaeomagnetic age: ~1.2 Myr old)⁸. In western Europe, hand-axes are found in sites ranging from 35 to ~500 kyr ago^{4,5,9}.

We are reporting Early Pleistocene (~0.9 Myr ago) and initial Middle Pleistocene (0.76 Myr ago) dates for hand-axes excavated from two western Mediterranean sites (Fig. 1). Solana del Zamborino (La Solana del Zamborino, 37.39° N, 3.11° W), long considered one of the younger Acheulian sites in the Iberian Peninsula¹⁰, and Estrecho del Quípar (Cueva Negra de Estrecho del Río Quípar, 38.04° N, 1.88° W) were previously assumed to be from the end of the Middle Pleistocene (~200 kyr ago)^{11,12}. More recent studies of micro-mammals indicated older ages within the Middle Pleistocene^{13,14}. However, our magnetostratigraphic analysis (Fig. 2) at Solana del Zamborino now indicates a date (~760 kyr ago) at the boundary between the Early and Middle Pleistocene, corresponding to the last palaeomagnetic polarity reversal (Matuyama to Brunhes chrons). At Estrecho del Quípar, our new date (~900 kyr ago) is Early Pleistocene, fully within the reverse polarity Matuyama chron. These new dates expose Acheulian lithic technology in the western Mediterranean before the end of the Early Pleistocene, comparable to that found in the eastern Mediterranean ('Ubeidiya)⁸.

Solana del Zamborino has been a reference for the terminal part of the Early Palaeolithic in Spain, with artefacts described as evolved Mode 2 (ref. 10) and Final Acheulian¹¹. Located in the Guadix Basin, a large western tributary of the internally drained Baza Basin, the Solana del Zamborino fossil quarry is an open-air site within a long (>150 m) stratigraphic sequence dominated by fluvial deposits derived from the high southern mountains (Sierra

Nevada >3,000 m). During the time of Solana del Zamborino, the fluvial deposition was displaced as an expanding palaeo-lake created palustrine (marsh) and littoral (shallow lake margin) environments. Excavations were made only during the mid-1970s, supplying an abundance of large-animal bones, tusks and charcoal (and other evidence of fire), along with several hand-axes and Palaeolithic tools made from quartzite, quartz and chert¹¹ (also see Supplementary Information).

The youthful age (~200 kyr old)¹¹ assumed for Solana del Zamborino was largely based on its well-developed Acheulian lithic typology. Such a young age contrasts with our continuing lithostratigraphy and palaeoclimate research in the region^{15–17}, which indicates a final, major lake-forming event near the end of the Early Pleistocene (starting ~800 kyr ago) and deposition terminating in the Baza Basin (~600 kyr ago). A specific question in chronology arose after a study at the Cúllar Baza-1 fossil quarry, 50 km to the east, where a detailed magnetostratigraphy¹⁶ found the Matuyama/Brunhes boundary just a few metres below the fossil/tool levels. Solana del Zamborino is in a similar palaeo-environmental setting, with similar micro-mammals (for example, *Arvicola cantianus*)¹³ and now has a similar magnetostratigraphy with the Matuyama/Brunhes boundary only a few metres below the fossil/tool-bearing levels. Although much older at 760 kyr, this new age is more compatible with our palaeoclimate-driven model in which precipitation was the primary control for palaeo-lake levels, directly influencing

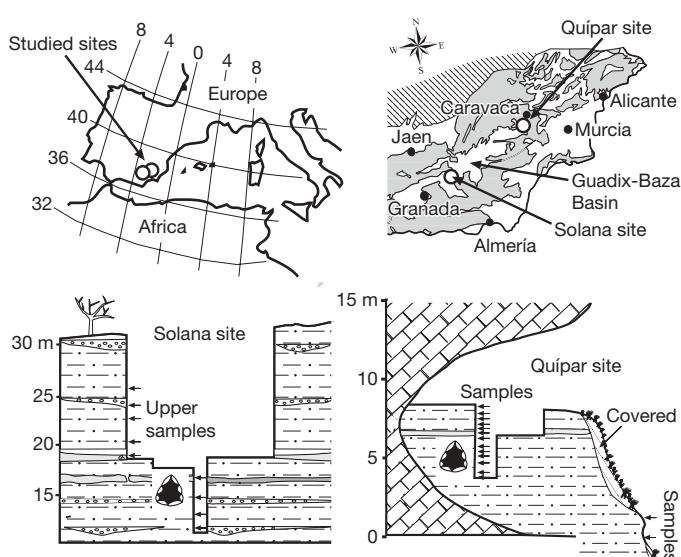


Figure 1 | Location maps and geological sketch. The two sites in the Iberian peninsula are La Solana del Zamborino and Cueva Negra del Estrecho del Río Quípar. Schematic cross-sections for each site indicate the levels with hand-axes (symbol) and palaeomagnetic samples (arrows). Four other palaeomagnetic samples were collected below the Solana quarry (not shown).

¹Berkeley Geochronology Center, 2455 Ridge Road, Berkeley, California 94709, USA.

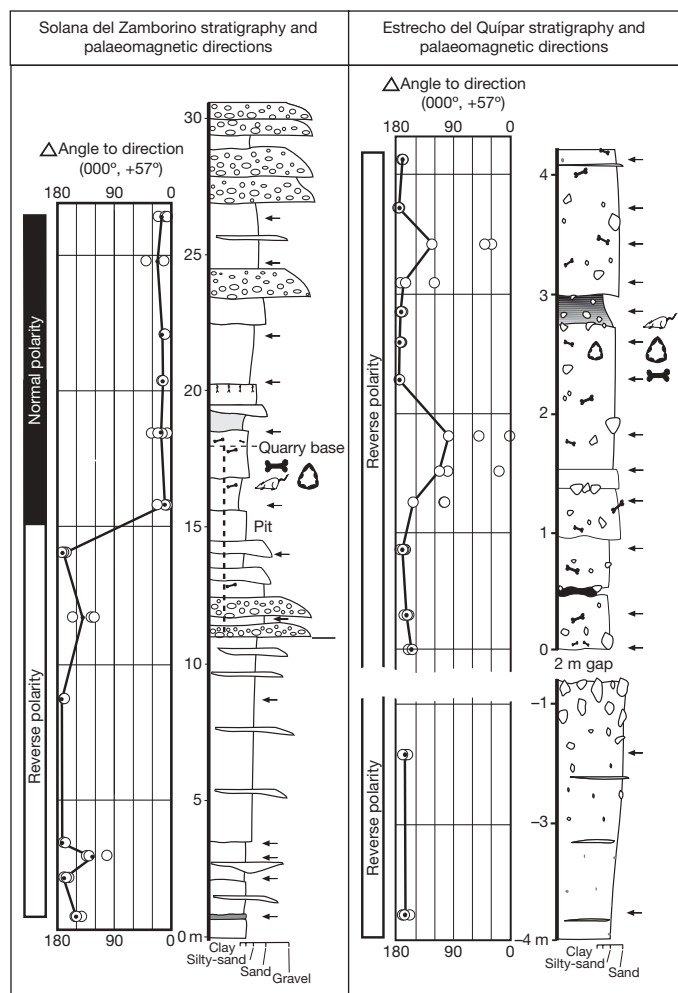


Figure 2 | Magnetostratigraphy and lithostratigraphy. The palaeomagnetic results from Solana del Zamborino and Estrecho del Quípar are shown with lithostratigraphic columns. Palaeomagnetic samples (arrows) are from vertical excavation faces, except the lower five samples at Solana del Zamborino and the lower two samples at Estrecho del Quípar. Palaeomagnetic remanence directions (Δ , from three specimens per level) are shown relative to the direction of the axial dipole field. Symbols: bone, large mammals; mouse, small mammals; broken cobble, lithic artefacts.

deposition and facies migration throughout the endorheic Guadix-Baza Basin^{15–17}. Both lakeside sites have similar mammals (Table 1), revealing an initial fauna of the Middle Pleistocene (Toringian biostratigraphic stage)¹⁶.

Below the artefact levels at Solana del Zamborino we found 14 m ($n = 7$ samples) of reverse polarity, whereas within and above these levels we found 12 m ($n = 6$) of normal polarity. Most of the section is composed of medium to coarse-grained sandstone, but for palaeomagnetic analysis we preferentially sampled finer-grained beds (siltstone and claystone). Solana del Zamborino's artefact layers are positioned immediately above the Matuyama/Brunhes polarity reversal. Using the sediment accumulation rate from the stratigraphically equivalent Cúllar Baza-1 section¹⁶, the tool layers would be 10–30 kyr after the polarity change, giving a date of 770–750 kyr ago.

Estrecho del Quípar is a rock shelter site on the northeastern margin of the Baza Basin near Caravaca de la Cruz (Murcia). Since 1990, annual excavations have generated an abundance of complex lithic artefacts and animal remains, along with hominin teeth, a hand-axe, and a large bifacial chopper¹⁴. Age estimates were initially Late Pleistocene¹² and more recently changed to Middle Pleistocene¹⁴. However, our analysis indicates reverse palaeomagnetic polarity throughout this 10-m accumulation of palaeosols.

Table 1 | Mammals from southeast Iberian sites

| Palaeomagnetic chronology | Pre-Jaramillo | | | Post-Jaramillo | | Brunhes | |
|---|---------------|----|------|----------------|------|---------|----|
| | VM | BL | FN-3 | CV | HU-1 | CB-1 | SZ |
| Sequences sites (Baza Basin) | | | | | | | |
| Cavern sites | | | | | | | |
| Micro-mammals | | | | | | | |
| <i>Castillomys crusafonti rivas</i> | X | X | X | X | X | — | — |
| <i>Microtus (Allophaiomys) cf. pliocaenicus</i> | X | X | X | — | — | — | — |
| <i>Mimomys savini</i> | — | — | X | — | X | X | — |
| <i>Microtus (Allophaiomys) chalinei</i> | — | — | X | X | — | X | — |
| <i>Pliomys episcopolis</i> | — | — | — | — | — | X | — |
| <i>Microtus (Iberomys) aff. huescarensis</i> | — | — | — | — | X | X | — |
| <i>Microtus (Iberomys) brecciansis</i> | — | — | — | — | X | — | X |
| <i>Arvicola cantianus</i> | — | — | — | — | — | X | X |
| Large mammals | | | | | | | |
| <i>Praeovibos</i> sp. | X | X | X | — | — | — | — |
| <i>Pachycrocuta brevirostris</i> | X | X | X | X | — | — | — |
| <i>Megantereon</i> sp. | X | X | X | X | — | — | — |
| <i>Ursus etruscus</i> | X | X | X | X | — | — | — |
| <i>Canis etruscus</i> | X | X | X | X | X | — | — |
| <i>Mammuthus meridionalis</i> | X | X | X | X | — | — | — |
| <i>Hippopotamus antiquus</i> | X | X | X | X | — | — | — |
| <i>Eucladoceros giulii</i> | X | X | X | — | — | — | — |
| <i>Megaloceros</i> sp. | — | — | — | X | X | X | — |
| <i>Homotherium</i> sp. | X | X | X | X | X | — | — |
| <i>Stephanorhinus etruscus</i> | X | X | X | X | X | X | — |
| <i>Equus altidens</i> | X | X | X | X | X | X | — |
| <i>Panthera gombaszoegensis</i> | — | — | — | X | X | — | — |
| <i>Elephas antiquus</i> | — | — | — | — | X | — | — |
| <i>Hippopotamus major</i> | — | — | — | — | X | — | — |
| <i>Bison</i> sp. | — | — | — | — | X | X | X |
| <i>Canis mosbachensis</i> | — | — | — | — | X | X | X |
| <i>Sus</i> sp. | — | — | — | — | X | X | X |
| <i>Dama 'nestii' vallonetensis</i> | — | — | — | — | X | — | — |
| <i>Ursus</i> sp. | — | — | — | — | X | — | — |
| <i>Macaca</i> sp. | — | — | — | — | X | X | X |
| <i>Crocota crocuta</i> | — | — | — | — | — | X | — |
| <i>Equus cf. sussenbornensis</i> | — | — | — | — | — | X | X |
| <i>Mammuthus trogontherii</i> | — | — | — | — | — | X | X |
| <i>Homo</i> sp. | X | X | — | X | — | X | — |
| Lithic artefacts | — | X | X | — | — | X | X |

Fossil mammals from Solana del Zamborino¹³ (SZ) and Estrecho del Quípar¹⁴ (EQ) compared with magnetostratigraphically calibrated sites in the Baza Basin^{15,16} and the cavern/mine site of Cueva Victoria, Murcia (CV)²⁹. The SZ fauna match the Cúllar Baza-1 (CB-1) site, which is also just above the Matuyama/Brunhes magnetochron boundary. The EQ fauna most closely correspond to the stratigraphically sequenced reference site Huéscar-1 (HU-1) in reverse polarity between the Jaramillo subchron and the Brunhes chron. Other site names: BL, Barranco León; FN-3, Fuentenueva-3; VM, Venta Micena. X, present at site; —, not present at site.

Therefore, the age must be Early Pleistocene, the most recent period dominated by reverse polarity (1.78–0.78 Myr ago)¹⁸.

Human fossils from the Estrecho del Quípar excavation include canine and premolar teeth with non-modern dimensions^{10,14}. The large bifacial tools (hand-axe and chopping tool) were fashioned from limestone cobbles and excavated ~1.5 m below the top of the lithified Pleistocene sediment¹⁴. Other lithic artefacts (80% chert) and abundant knapping debris are ubiquitous to the current limit of archaeological excavation (depth ~4.5 m). Lithic artefacts include small disc-cores and flakes, some from prepared striking platforms and some with centripetal and recurrent flaking, or with abrupt or semi-abrupt edge retouch¹⁴. The lithic artefacts from Estrecho del Quípar reveal some of the diversity in core-reduction techniques already in use during the Early Pleistocene in southern Iberia.

The Estrecho del Quípar deposit consists of ~10 m of fine-grained fossiliferous sediment infilling one of the relict weathering cavities (tafoni) which are common features in the massive limestone cliffs in this part of the Quípar Valley. An upper 4.2 m archaeological section was densely sampled for palaeomagnetic analysis ($n = 13$). A lower ~2 m exploratory section was sampled ($n = 2$) from outcrop, outside the present roof of the rock shelter. The entire deposit has both alluvial and soil characteristics, which we interpret as a sequence of

accumulating immature palaeosols. All beds have pedogenic micro-fabric with soil plasma incorporating variable amounts of detrital grains (rounded, sand-size limestone and quartz), tafoni weathering flakes (carbonate cemented, 1–10 mm), bone fragments plus chert/limestone artefacts (sharp-edged), and occasionally large angular limestone fragments (rock falls, to boulder size).

Reverse palaeomagnetic polarity was found throughout the Estrecho del Quípar sequence including a sample within 10 cm of the top of the deposit (Fig. 2). Therefore, the entire sequence, with >8 m of reversely magnetized palaeosols, is more than 780 kyr old, before the start of the Brunhes normal polarity chron (C1n). The youngest magneto-chronological assignment is to the late Matuyama subchron (C1r.1r) in the age range 990–780 kyr ago¹⁸. Comparing the Estrecho del Quípar micro-mammals (for example, *Microtus huescarensis*) to the chronostratigraphic sequence from the Baza Basin (Table 1) provides a close correspondence to Huéscar-1, a reverse polarity fossil-quarry. Huéscar-1 is ~10 m below the Matuyama/Brunhes polarity boundary (~900 kyr ago)¹⁶. A necessarily approximate date of ~900 kyr ago for Estrecho del Quípar should be used until a zone of normal polarity lower in the sequence might be delimited, or other chronological refinements develop.

There are other possible, although less likely, correlations for the polarity zones being reported for these two hand-axe sites. These would produce older dates of 1.06 Myr ago for Solana del Zamborino, and ~1.2 Myr ago for Estrecho del Quípar. Such older correlations would conflict with the reference chronostratigraphy (see Table 1) being developed from long overlapping stratigraphic sections within the adjacent Baza Basin^{15–17}. For example, *M. huescarensis*, found at Estrecho del Quípar, has not been found in reverse polarity sites in the stratigraphic sequence, such as Venta Micena or Fuentenueva-3 (1.3–1.2 Myr ago)¹⁵, until much higher at Huéscar-1 (~0.9 Myr ago)¹⁶.

This report confirms the utility of long reference chronostratigraphies as age constraints for more isolated sites such as caverns and short stratigraphic sections. The Baza reference chronostratigraphy^{15,16} that we are using had recently clarified the chronology of sites already in use as micro-mammal calibration sites for western Europe. This recalibration generated a clearer biochronological view at the boundaries of the Pliocene to Pleistocene^{15,19}, and the Early to Middle Pleistocene¹⁶. Of specific interest for the chronology of other hand-axe sites in Europe was the early record of *Arvicola cantianus* immediately above the Matuyama/Brunhes polarity boundary (~0.75 Myr ago)¹⁶. This antiquity is confirmed by the polarity zonation at Solana del Zamborino. Unfortunately, deposition in the Baza Basin ended by ~0.5 Myr ago¹⁶, leaving the biochronology of the last half of the Middle Pleistocene to be developed elsewhere.

The large change in dates for the two hand-axe sites in this report reiterates the warning against a simple connection between lithic typology and chronology⁴. The dates accepted for the numerous, widely dispersed archaeological sites in Europe should be critically evaluated and linked to developing chronostratigraphic sequences whenever available (see, for example, ref. 20) or linked to sites that have materials amenable for radiometric analysis (see, for example, ref. 21). The Middle Pleistocene remains a challenging period for chronological studies, lacking the worldwide palaeomagnetic reversals that are useful in identifying and subdividing the Early Pleistocene¹⁸.

Overall, the Early Pleistocene was a time with widespread distribution of Palaeolithic bifacial technology outside of Africa, extending across the eastern⁷, southern⁸, and now the western limits of temperate Eurasia. Our palaeomagnetic analysis indicates an older than expected^{4,5} Acheulian chronology in the western Mediterranean, and when these are combined with the Oldowan (uni-facial) lithic technology²² in the Baza Basin sites at Orce, Spain (palaeomagnetic age: ~1.3 Myr old)¹⁵, a hominin presence^{23,24} in southwestern Europe is indicated for much of the Early Pleistocene (Fig. 3). This developing outline of a longer European hominin chronology also suggests that the barrier between Africa and Europe was permeable. Establishing

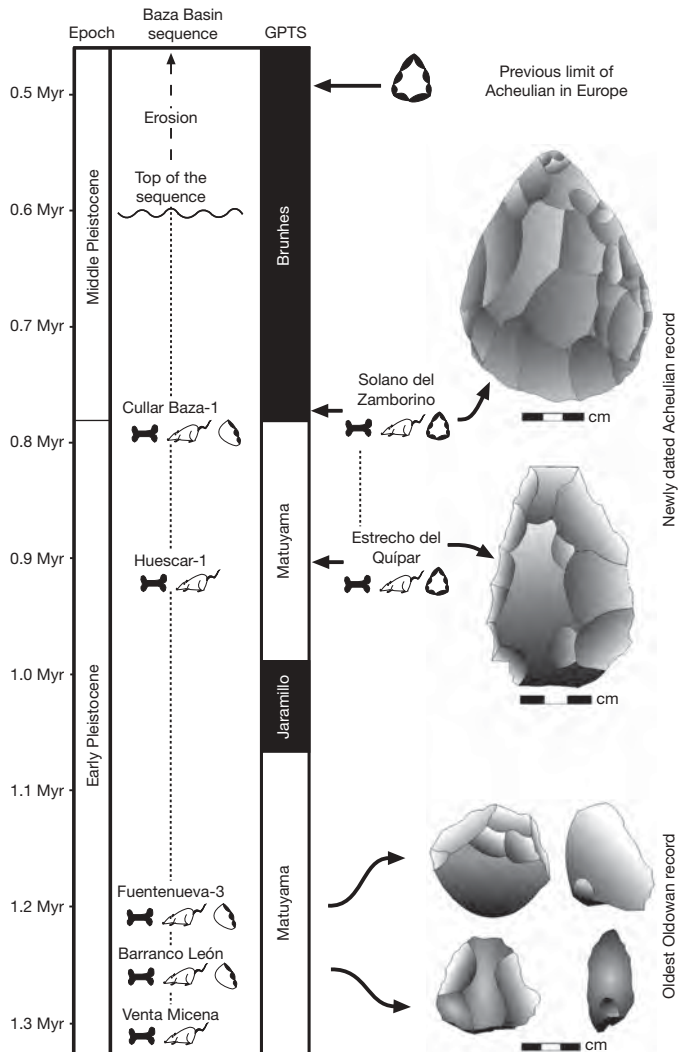


Figure 3 | Magnetochronology of Palaeolithic sites. The chronology of Early Palaeolithic sites in southern Iberia referenced to the combined magnetochronology and biochronology from lithostratigraphic sequences in the Baza Basin^{13–16}. Note the mid-Middle Pleistocene age of the assumed oldest Acheulian sites in Europe^{4,5}. GPTS, geomagnetic polarity timescale (see ref. 18). Drawings of Palaeolithic tools modified from refs 11, 14 and 22; for other symbols see Fig. 2. Scale bar divisions, 1 cm.

secure chronologies for early hominin sites will continue to increase our understanding of human prehistory in Eurasia and its relation to Africa^{9,24–26}.

METHODS SUMMARY

Blocks (~12 cm on a side) were oriented while still attached to the substrate using magnetic compass and inclinometer. Each sample was sawn without water into three specimen cubes (12 cm³ each), and the surfaces were sanded (to remove metal contamination) and cleaned with compressed air. Storage, measurements and laboratory experiments were made inside a room-sized magnetostatic shield (<350 nT), equipped with a superconducting rock magnetometer (sensitivity 10⁻¹² Am²), an inline alternating field demagnetizer (3-axis, static), a non-inductive furnace (residual field < 3 nT), and a magnetic susceptibility meter (inducing field 100 μT).

The remanent magnetism preserved in individual sedimentary strata were measured and then assembled in stratigraphic position to form palaeomagnetic polarity zones. Of specific interest is finding the last complete polarity reversal in the Earth's magnetic field, 780 kyr ago²⁷. This global chronological marker is used to approximate the boundary between the Early and Middle Pleistocene²⁸. Most samples from these hand-axe sites have a well-preserved remanent magnetism. Initial demagnetization experiments removed the viscous and weathering magnetizations, which are directed parallel to the modern magnetic field (north and down). Further thermal demagnetization exposed the characteristic

magnetization that has either reverse or normal polarity directions (see Supplementary Information for details).

Received 27 November 2008; accepted 16 June 2009.

1. Asfaw, B. *et al.* The earliest Acheulean from Konso-Gardula. *Nature* **360**, 732–734 (1992).
2. Isaacs, G., Li, C. & Curtis, G. H. Age of early Acheulean industries from the Peninj Group, Tanzania. *Nature* **249**, 624–627 (1974).
3. Chazan, M. *et al.* Radiometric dating of the Earlier Stone Age sequence in Excavation I at Wonderwerk Cave, South Africa: preliminary results. *J. Hum. Evol.* **55**, 1–11 (2008).
4. Monnier, G. F. The Lower/Middle Palaeolithic periodization in Western Europe: an evaluation. *Curr. Anthropol.* **47**, 709–744 (2006).
5. Santonja, M. & Villa, P. in *Axe Age. Acheulian Toolmaking from Quarry to Discard* (eds Goren-Inbar, N. & Dibble, H. L.) 429–478 (Equinox, 2006).
6. Goren-Inbar, N. *et al.* Pleistocene milestones on the Out-of-Africa Corridor at Geshen Benot Ya'aqov, Israel. *Science* **289**, 944–947 (2000).
7. Yamei, H. *et al.* Mid-Pleistocene Acheulean-like stone technology of the Bose Basin, south China. *Science* **287**, 1622–1626 (2000).
8. Bar-Yosef, O. & Goren-Inbar, N. The lithic assemblages of 'Ubeidiya: a Lower Palaeolithic site in the Jordan Valley. *Monogr. Inst. Archaeol. Qedem* 34 (Hebrew University of Jerusalem, 1993).
9. Bermúdez de Castro, J. M. *et al.* The Atapuerca sites and their contribution to the knowledge of human evolution in Europe. *Evol. Anthropol.* **13**, 25–41 (2004).
10. Aguirre, E. *Homo hispánico* (Editorial Espasa Calpe, 2008).
11. Botella, M. C., Vera, J. A. & de Porta, J. El yacimiento achelense de la "Solana del Zamborino", Fonelas (Granada). Primera campaña de excavaciones. *Cuad. Prehist. Univ. Granada* **1**, 1–45 (1975).
12. Walker, M. J. in *A Very Remote Period Indeed. Papers on the Palaeolithic presented to Derek Roe* (eds Milliken, S. & Cook, J.) 153–159 (Oxbow Books, 2001).
13. Ruiz Bustos, A. in *The Hominids and their Environment in the Middle and Lower Pleistocene of Eurasia* (eds Gibert, J., Sánchez, F., Gibert, L. & Ribot, F.) 153–174 (Museo de Prehistoria y Paleontología de Orce, 1995).
14. Walker, M. J. *et al.* Cueva Negra del Estrecho del Río Quípar (Murcia, southeast Spain): An Acheulian and Levallois-Mousteroid assemblage of Palaeolithic artifacts excavated in a Middle Pleistocene faunal context with hominin skeletal remains. *Eurasian Prehistory* **4**, 3–43 (2006).
15. Scott, G. R., Gibert, L. & Gibert, J. Magnetostratigraphy of the Orce region (Baza Basin), SE Spain: new chronologies for Early Pleistocene faunas and hominin occupation sites. *Quat. Sci. Rev.* **26**, 415–435 (2007).
16. Gibert, L., Scott, G., Martin, R. & Gibert, J. The Early to Middle Pleistocene boundary in the Baza Basin (Spain). *Quat. Sci. Rev.* **26**, 2067–2089 (2007).
17. Gibert, L., Ortí, F. & Rosell, L. Plio-Pleistocene lacustrine evaporites of the Baza Basin (Betic Chain, SE Spain). *Sed. Geol.* **200**, 89–116 (2007).
18. Gradstein, F. M. *et al.* *A Geological Time Scale 2004* (Columbia Univ. Press, 2004).
19. Gibert, L., Scott, G. & Ferrández-Cañadell, C. Evaluation of the Olduvai subchron in the Orce ravine (SE Spain). Implications for Plio-Pleistocene mammal biostratigraphy and the age of Orce archaeological sites. *Quat. Sci. Rev.* **25**, 507–525 (2006).
20. Parfitt, S. A. *et al.* The earliest record of human activity in northern Europe. *Nature* **438**, 1008–1012 (2005).
21. Coltorti, M. *et al.* New $^{40}\text{Ar}/^{39}\text{Ar}$, stratigraphic and palaeoclimatic data on the Isernia La Pineta Lower Palaeolithic site, Molise, Italy. *Quat. Int.* **131**, 11–22 (2005).
22. Gibert, J., Gibert, L., Ferrández-Cañadell, C., Iglesias, A. & González, F. in *The Human Evolution Source Book* (eds Ciochon, R. L. & Fleagle, J. G.) 327–335 (Pearson Prentice Hall, 2007).
23. Borja, C., García-Pacheco, M., Olivares, E. G., Scheuenstuhl, G. & Lowenstein, J. M. Immunospecificity of albumin detected in 1.6 million-year-old fossils from Venta Micena in Orce, Granada, Spain. *Am. J. Phys. Anthropol.* **103**, 433–441 (1997).
24. Gibert, J. *El Hombre de Orce: los Hominidos que Llegaron del sur* (Editorial Almuzara, 2004).
25. Dennell, R. Dispersal and colonisation, long and short chronologies: how continuous is the Early Pleistocene record for hominids outside East Africa? *J. Hum. Evol.* **45**, 421–440 (2003).
26. Anton, S. C. & Swisher, C. C. Early dispersals of *Homo* from Africa. *Annu. Rev. Anthropol.* **33**, 271–296 (2004).
27. Coe, R. S., Singer, B. S., Pringle, M. S. & Zhao, X. Matuyama–Brunhes reversal and Kamikatsura event on Maui: Palaeomagnetic directions, $^{40}\text{Ar}/^{39}\text{Ar}$ ages and implications. *Earth Planet. Sci. Lett.* **222**, 667–684 (2004).
28. Ciaranfi, N., van Kolfschoten, T. & Coltorti, M. The Plio-Pleistocene boundary and the Lower/Middle Pleistocene transition: type areas and sections—an introduction. *Quat. Int.* **131**, 1–3 (2005).
29. Gibert, J. *et al.* Cueva Victoria: Geología, Paleontología, restos humanos y edades. *Mem. Arqueología Región Murcia* **14**, 37–62 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Partial funding and support was from the Fullbright Scholar Program (L.G.), Fundación Española para la Ciencia y la Tecnología (L.G.), project 05584/ARQ/07 of the Fundación Séneca (M. Walker), project CGL2005-05337BTE of the Dirección General de Investigación (L. Rosell), and the Earthwatch Institute (L.G.). For discussions, laboratory, and field assistance we thank J. Gibert, M. Walker, R. Martin, C. Ferrández, J. Butterworth, L. Hinton, M. Lería, A. López-Jiménez, M. López-Martínez, S. Matson, J. Ortega, F. Ribot, A. Ruiz-Bustos and L. Smeenk. We also thank the Museo Arqueológico de Vélez Rubio (Almería) for access to their collection from Solana del Zamborino, R. Gómez Torres for access to the Solana del Zamborino site, and M. Walker for access to fossils/artefacts and sampling at Estrecho del Quípar. We acknowledge the late J. Gibert Clols for encouraging the development of this research and providing the large mammal fauna identifications.

Author Contributions The authors of this paper contributed equally. G.R.S. directed the palaeomagnetic measurements and assisted in field sampling, and L.G. directed the palaeomagnetic field sampling and assisted in laboratory measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.G. (lgibert@bgc.org).

LETTERS

iPS cells produce viable mice through tetraploid complementation

Xiao-yang Zhao^{1,2*}, Wei Li^{1,2*}, Zhuo Lv^{1,2*}, Lei Liu¹, Man Tong^{1,2}, Tang Hai¹, Jie Hao^{1,2}, Chang-long Guo^{1,2}, Qing-wen Ma³, Liu Wang¹, Fanyi Zeng^{3,4} & Qi Zhou¹

Since the initial description of induced pluripotent stem (iPS) cells created by forced expression of four transcription factors in mouse fibroblasts, the technique has been used to generate embryonic stem (ES)-cell-like pluripotent cells from a variety of cell types in other species, including primates and rat^{1–6}. It has become a popular means to reprogram somatic genomes into an embryonic-like pluripotent state, and a preferred alternative to somatic-cell nuclear transfer and somatic-cell fusion with ES cells^{7,8}. However, iPS cell reprogramming remains slow and inefficient. Notably, no live animals have been produced by the most stringent tetraploid complementation assay, indicative of a failure to create fully pluripotent cells. Here we report the generation of several iPS cell lines that are capable of generating viable, fertile live-born progeny by tetraploid complementation. These iPS cells maintain a pluripotent potential that is very close to ES cells generated from *in vivo* or nuclear transfer embryos. We demonstrate the practicality of using iPS cells as useful tools for the characterization of cellular reprogramming and developmental potency, and confirm that iPS cells can attain true pluripotency that is similar to that of ES cells.

Several studies have shown that iPS cells resemble ES cells in morphology, gene expression profiles, and the epigenetic status of several pluripotency markers^{9,10}. Notably, iPS cells have maintained their ability to generate chimaeras that are competent for germline transmission^{9–14}. Their therapeutic potential is thought to be similar to that of ES cells in several disease mouse models^{15,16}. Despite the promise this technique holds for the development of patient-specific cell therapy, considerable technical barriers remain that must be overcome for application as a therapeutic intervention. To reduce the potential oncogenic effect of genomic integration of viral vectors, different combinations or reduced numbers of factors have been introduced, as well as the use of alternative delivery methods and agents^{11,13,17–21}. However, one of the remaining technical hurdles in the reprogramming process is the slow and inefficient induction of iPS cells, especially those with a complete developmental potential capable of producing live-born offspring. Tetraploid complementation is considered the most stringent test for pluripotency and developmental potency, as any viable live-born animals resulting from the injection of diploid ES or iPS cells to create the tetraploid (4N) embryos (blastocysts) will be from the diploid donor cells. The tetraploid host blastocyst primarily contributes to the extra-embryonic lineages and not to the embryo proper^{22,23}. Unlike pluripotent ES cells generated from *in vivo* or somatic-cell nuclear transfer (SCNT)-produced embryos, all previous studies have failed to produce viable offspring from iPS cells through tetraploid complementation^{9,13,24}.

Here we report the generation of 37 iPS cell lines that not only express the correct pluripotency markers by *in vitro* assays, but many are also able to produce chimaeric mice with germline transmission by blastocyst injection. Three of these lines produced viable, live-born offspring by tetraploid complementation.

We infected mouse embryonic fibroblast (MEF) cells expressing an Oct4 (also known as Pou5f1)-enhanced green fluorescent protein (GFP) reporter with the four 'Yamanaka factors'^{1,25}: pMXs-Oct4, Sox2, Klf4 and c-Myc, and cultured the cells as previously reported²⁵ with modifications such that knockout serum replacement (KOSR)^{26,27} was substituted for fetal bovine serum (FBS) for iPS cell derivation. At day 10, alkaline phosphatase⁺ and GFP⁺ colonies were detected and iPS cell lines were derived at days 14, 20 and 36 (Fig. 1a). From several experimental runs, 31 GFP⁺ colonies were obtained and yielded stable cell lines (Supplementary Table 1). In addition to the Oct4–GFP MEF cells from a B6D2F1 genetic background, six iPS cell lines from MEF cells with a C57 × 129S2 background were also generated (IP14D-101–106). Because our laboratory does not maintain any ES or SCNT-ES cell lines of this genetic background, these iPS cells could only result from induced fibroblasts (Table 1).

Pluripotency markers such as Oct4, Nanog and SSEA1 were expressed (Fig. 1b), and the cell lines were predominantly diploid with normal 40 chromosome karyotypes (Supplementary Fig. 1). Bisulphite sequencing showed demethylation of *Oct4* and *Nanog* promoters in three iPS cell lines tested compared to the parental Oct4–GFP MEF cells, a pattern similar to that from normal ES cells (Supplementary Fig. 2) reflecting the epigenetic remodelling that occurred during reprogramming. Teratomas with all three germ layers determined by histological analyses were observed from iPS cells injected into severe combined immunodeficient (SCID) mice (Fig. 1c), indicating a considerable degree of pluripotency. Quantitative reverse transcription–polymerase chain reaction (RT–PCR; Supplementary Fig. 3) demonstrated that the transgenes were almost completely silenced in the iPS cell lines tested (IP36D, IP20D and IP14D), similar to previous reports^{1,9} suggesting that the maintenance of iPS cell lines mainly requires the endogenous genes of these four factors.

To test pluripotency further, we randomly selected one or two cell lines from each of the experimental runs, and injected them into normal CD-1 blastocysts that were transferred to CD-1 pseudopregnant recipient females. As shown in Table 1, IP36D and IP20D cell lines produced many live mice with 5–80% chimaerism, qualitatively assessed by coat colour. Germline transmission was noted in two of the four lines. Five of the thirteen IP14D iPS cell lines (four with B6D2F1 and one with a C57 × 129S2 background) were randomly selected, and all resulted in live chimaeric mice with 10–95% chimaerism. From

¹State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. ²Graduate School of Chinese Academy of Sciences, Beijing 100049, China. ³Shanghai Institute of Medical Genetics, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai 200040, China. ⁴Institute of Medical Science, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China.

*These authors contributed equally to this work.

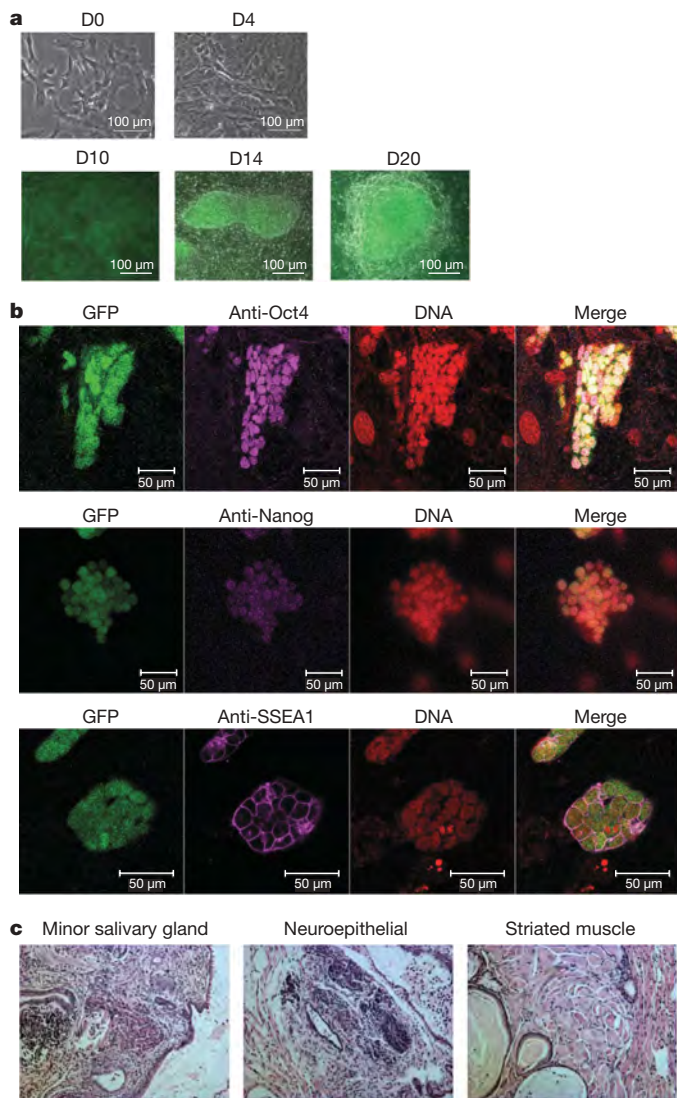


Figure 1 | Characterization of the iPS cells generated in 20% knockout serum replacement culture systems. **a**, Cell morphology. Top, Oct4–GFP MEF cells at day (D) 0 (before viral infection) and day 4 (just before replating onto MEF feeder cells). Bottom, morphology of GFP⁺ cells at days 10, 14 and 20. **b**, Immunostaining for pluripotency markers in Oct4–GFP iPS cells. Positive GFP fluorescence (green) and Oct4, Nanog and SSEA1 (purple) were observed. DNA was stained by propidium iodide (red). Shown are examples from the IP20D-3 line. The other two cell lines analysed also gave the same pattern. **c**, Teratoma formation. Sections were stained with haematoxylin and eosin. Shown are tissues representative of all three germ layers: minor salivary gland (endoderm), neuroepithelial (ectoderm) and striated muscle (mesoderm). Original magnification, $\times 100$.

this source, three of the five lines were germline-transmitted to the next generation (Fig. 2). Thus, our iPS cell lines are competent for germline transmission with higher efficiency in the IP14D group compared to IP20D and IP36D.

Until now, the developmental potential of iPS cells could not be fully evaluated because embryos generated from iPS cells using tetraploid complementation ('iPS 4N-comp') were viable only into the later stages of gestation^{9,13,24}, demonstrating that complete pluripotency had not been attained. When we injected our iPS cells into tetraploid CD-1 blastocysts (white coat), we observed complete development potential resulting in the birth of live pups. Different iPS cell lines were not equally successful in producing viable offspring, although all were competent for germline transmission through chimaeric mice. Cell lines IP36D-3, IP20D-3 and IP20D-19 showed early termination of fetal development at embryonic day (E) 13.5 and E15.5 (Table 1). There seems to be a developmental advantage for the cell line derived at day 20 compared to day 36 (8.8% of IP20D-19 versus 1.7% of IP36D-3 reached E13.5). One IP20D-3 embryo developed normally until E15.5 (Table 1). Figure 2a and b show iPS 4N-comp embryos with normal embryonic morphology collected at E9.5 and E13.5. In contrast, when we injected cells from IP14D-1 (derived from GFP⁺ colonies at day 14) into tetraploid blastocysts, 22 live-born pups were obtained (3.5%), as well as four embryos (0.6%) that developed to E17.5. Two other IP14D lines with B6D2F1 and C57 \times 129S2 backgrounds (IP14D-6 and IP14D-101, respectively) were able to produce live iPS 4N-comp pups. The rate of iPS 4N-comp animal production is similar to the rate from normal ES cells in our laboratory (Table 1). The IP14D 4N-comp pups have survived from 2 days to almost 9 months so far. The black mouse in Fig. 2f represents one of the live iPS 4N-comp mice at 15 weeks. Note the uniform black coat of this 'all-iPS' mouse, resembling the coat colour of the original line of the Oct4–GFP MEF cells (B6D2F1) from which it was derived.

To identify clearly the lineage of the cells and the 4N-comp mice, simple sequence length polymorphism (SSLP) analyses followed by PCR for various marker genes were performed. The iPS 4N-comp mice differed from the CD-1 mice at markers consistent with a B6D2F1 lineage or the C57 \times 129S2 origin, but had exactly the same profiles as the iPS cell lines from which they were derived (Fig. 2c). This was confirmed by PCR analysis of unique microsatellites (Fig. 2d). The iPS cell lines, 2N chimaeric mouse and offspring, iPS 4N-comp mice and their progeny mated to CD-1 females all yielded the expected inheritance patterns, confirming that the IP14D-1 lineage of the iPS 4N-comp adults is distinct from the CD-1 mouse. Furthermore, Southern blot analysis and reverse PCR results independently confirmed that the viral integration patterns in the iPS 4N-comp mice are identical to the three iPS cell lines from which they originated (Supplementary Figs 4 and 5).

To study the embryonic development of offspring from iPS 4N-comp mice, a 7-week-old male mouse carrying Oct4–GFP was mated to a CD-1 female. The iPS mouse was capable of impregnating the dam and producing viable preimplantation stage embryos with normal morphology and without developmental delay. GFP⁺

Table 1 | Developmental efficiency of embryos produced

| Cell line | 2N injection | | | | 4N injection (%) | | | |
|----------------|----------------------|----------------|----------------|-----------|----------------------|---------------------------------|---------------------------------|-----------|
| | Injected blastocysts | Live chimaeras | Chimaerism (%) | Germ line | Injected blastocysts | Embryos arrested at E10.5–E13.5 | Embryos arrested at E15.5–E17.5 | Live pups |
| IP14D-1 | 135 | 6 | 20–95 | Yes | 624 | 7 (1.1) | 4 (0.6) | 22 (3.5) |
| IP14D-6 | 97 | 10 | 20–90 | Yes | 43 | — | — | 1 (2.3) |
| IP14D-101 | 360 | 18 | 10–90 | Yes | 181 | 15 (8.3) | — | 4 (2.2) |
| IP20D-3 | 236 | 14 | 20–80 | Yes | 204 | 6 (2.9) | 2 (1.0) | 0 |
| IP20D-19 | 48 | 4 | 10–70 | Yes | 273 | 24 (8.8) | — | 0 |
| IP36D-3 | 233 | 6 | 5–70 | Not yet | 229 | 4 (1.7) | 0 | 0 |
| Control (ES1) | NT | — | — | — | 100 | — | — | 3 (3.0) |
| Control (CL11) | NT | — | — | — | 51 | 5 (9.8) | — | 1 (2.0) |

Embryos were produced by blastocyst injection (2N) and tetraploid complementation (4N). IP14D-101 has a genetic background of C57 \times 129S2, whereas the rest are from C57 \times DBA F₁ (B6D2F1). NT, not tested.

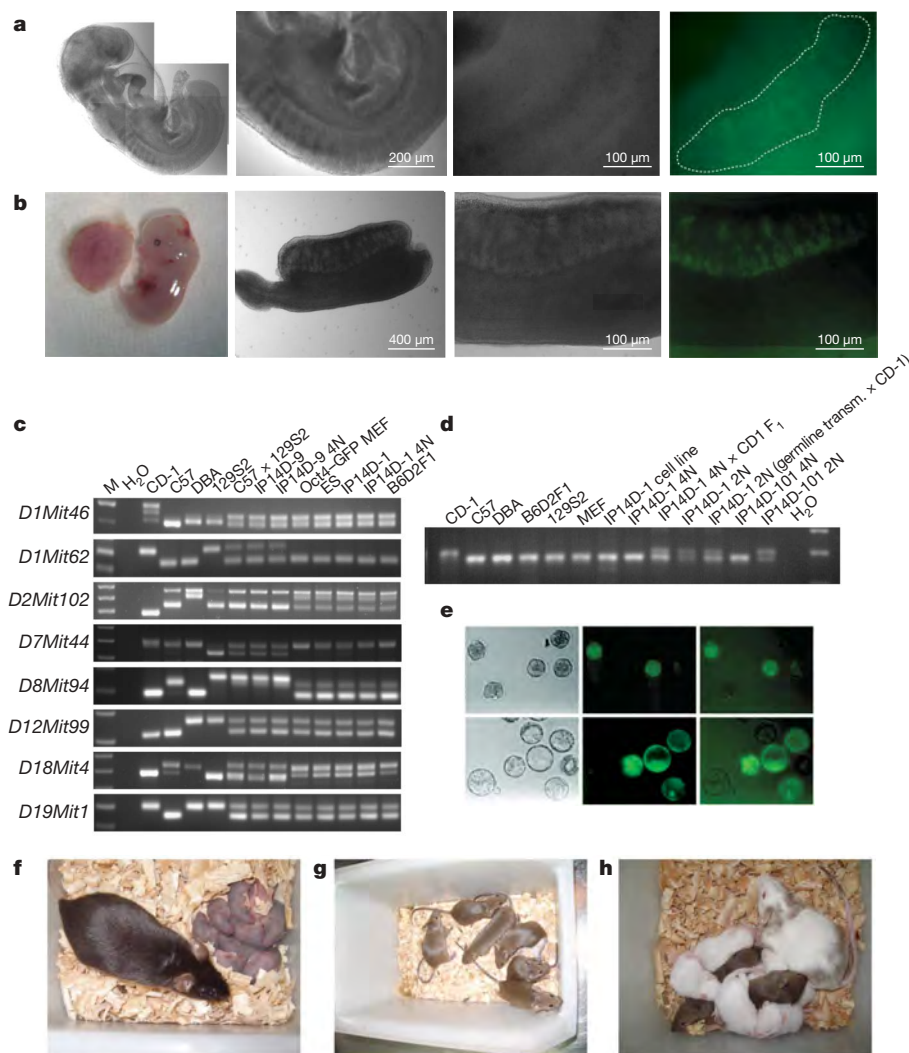


Figure 2 | In vivo developmental potential of iPS cell lines generated by tetraploid complementation. Ten to fifteen iPS cells (B6D2F1, black coat colour) were injected into each CD-1 (white coat colour) tetraploid blastocyst electrofused at the two-cell stage, and then transferred into pseudopregnant female recipients. Embryos derived from tetraploid blastocyst injection were dissected on E9.5, E13.5 and on the day of birth (E19.5). **a**, A live E9.5 IP36D-3-derived embryo (left) with 18 somites (middle left). Primordial germ cells migrate along hind-gut mesentery (middle right). Oct4-GFP-positive migrating primordial germ cells are shown (right). **b**, A live E13.5 embryo (left) with placenta derived from IP20D-3. Male gonad and mesonephros of fetus (middle left) are shown. Oct4-GFP is positive (right) in male gonadal germ cells in fetal testes (middle right). **c**, SSAP analysis for lineage identification covers markers from different chromosomes, and the 4N mice showed a polymorphic pattern similar to that from the parental MEF cells originating from B6D2F1 or C57 × 129S2 chimaeras, and different from the CD-1 tetraploid

embryos were observed at the morula and blastocyst stages, coinciding with the timing of Oct4 expression (Fig. 2e). All the progeny presented with a uniform brown coat colour typical of matings between a black B6D2F1 (that is, iPS 4N-comp) and a white CD-1 mouse (Fig. 2g). This is in contrast to the progeny obtained from the mating of the germline transmittable chimaeric mouse (2N) produced by the same iPS cells by blastocyst injection into a CD-1 mouse (Fig. 2h). This latter process gives rise to a combination of brown- and white-coated offspring. This further demonstrated the viability and normal productivity of the iPS 4N-comp mice, and reflects the true pluripotency of the iPS lines that are capable of producing live iPS 4N-comp mice.

Global gene expression in iPS cell lines that are capable of producing 4N-comp or 2N germline transmittable mice was analysed as

blastocyst donor. M denotes molecular mass marker. **d**, Microsatellite markers detect differences between CD-1 and parental patterns of the inbred C57 and DBA or a hybrid C57 × 129S2 strain, differentiating DNA from the iPS cell lines and the corresponding 4N-comp mouse, from a 2N chimaeric mouse or the 4N-comp × CD-1 F₁ offspring. **e**, Oct4-GFP positive morula (top) and blastocyst (bottom) flushed from CD-1 mated with an IP14D-1 4N mouse. Phase (left), fluorescence (middle) and merged (right) images are shown. Original magnification, ×200. **f**, A fifteen-week-old iPS-tetraploid male mouse derived from IP14D-1, with a uniformly black coat of the B6D2F1 strain from which Oct4-GFP MEF cells originate. The pups lying next to it are its F₁ progeny from its mating to a CD-1 dam. **g**, The F₁ pups shown in **f** developed a uniform brown coat, typical of a mating between B6D2F1 (that is, iPS 4N-comp) and CD-1 mice. **h**, A chimaeric mouse (2N) produced from blastocyst injection with IP14D-1 cells. When the mouse is mated to a CD-1 mouse, the progeny are a combination of brown and white coat colours.

reported previously²⁸ (Fig. 3). The expression of pluripotency markers (*Oct4*, *Nanog* and *Sox2*) as well as other regulators was similar between the iPS and the ES cells, distinct from that of the MEF cells (Fig. 3a and Supplementary Fig. 6). Hierarchical clustering analysis grouped all of the iPS cells with the ES cells, and not with the MEF cells (Fig. 3b). Clearly, the global expression pattern of 4N-comp iPS cells resembles ES cells that are tetraploid competent—an assay considered to be the gold standard for true pluripotency.

We generated 37 iPS lines that demonstrated ES-like characteristics and enhanced developmental potentials through retroviral infection with the four Yamanaka factors. Three iPS lines are capable of generating 4N-comp mice; so far 27 have been born, some of which have matured and are reproductively competent. The reprogramming of these iPS cells should be further investigated to identify

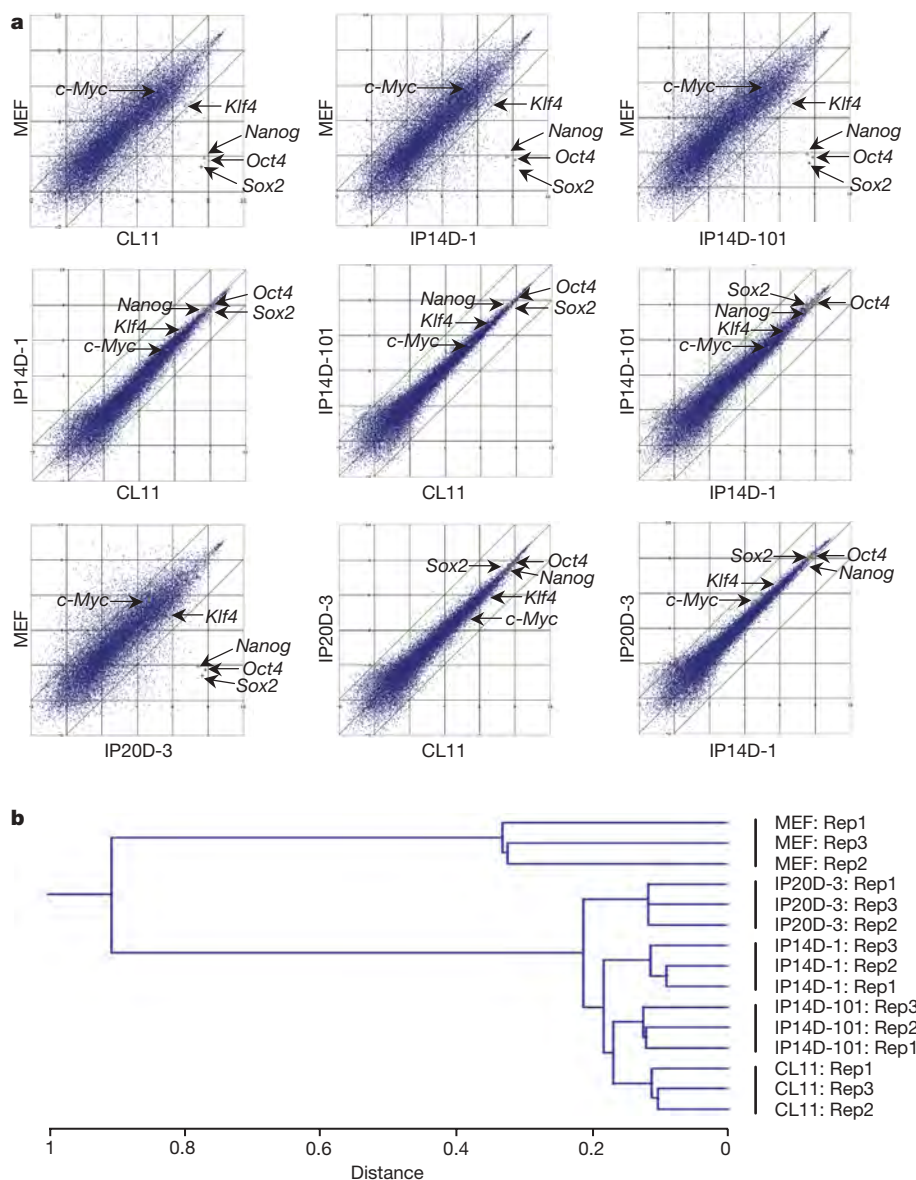


Figure 3 | Global gene expression analysis of iPS cell lines competent for tetraploid complementation. **a**, Scatter-plots compare the expression values for all probe sets from genome-wide transcription profiling derived from ES (CL11), IP14D-1 and IP14D-101 cell lines that are capable of 4N-comp, or IP20D-3 that is a germline transmittable iPS line, or the parental Oct4–GFP

MEF cells that produced IP14D-1 and IP20D-3. Parallel diagonal lines indicate twofold expression difference thresholds. **b**, Hierarchical clustering of microarray data for differentially expressed genes from these various lines. Tree branch distance scale = $1 - r$ in which r is Pearson's correlation coefficient. Rep, replicate.

factors critical for determining the higher developmental capacity of iPS cell lines derived from clones picked earlier in the reprogramming process, perhaps at time point(s) critical for epigenetic regulation. It would be interesting to determine whether iPS cells produced from direct reprogramming use different mechanisms for altering the genome compared to SCNT. Our result provides the first, to our knowledge, definitive evidence that iPS cells are truly pluripotent, a characteristic that is critical for important applications such as therapeutic interventions.

METHODS SUMMARY

iPS cells were generated from MEF cells with different genetic background as described¹, with the exception of culturing in 20% knockout serum replacement (KOSR) (Gibco) instead of FBS. Thirty-seven colonies were obtained for derivation, and all yielded stable cell lines. iPS cell lines derived at different time points were further confirmed for pluripotent properties by alkaline phosphatase staining and immunofluorescent analysis of SSEA1 (Chemicon), Oct4 (Santa Cruz) and Nanog (Chemicon). Teratoma assays were performed by injecting iPS cells into the subcutaneous flanks of SCID mice, followed by histological examination

of the tumours 4–5 weeks later. Bisulphite genomic sequencing, karyotype analysis, SSLP analysis, reverse PCR, and Southern blot analysis were performed. Tetraploid embryo complementation was carried out as described by injecting iPS cells (B6D2F1, black coat origin) into CD-1 tetraploid embryos (white coat colour), and the embryos derived from tetraploid blastocyst injection (4N) were dissected in handling medium on E9.5, E13.5 and the day of birth (E19.5), and analysed for morphology and developmental competency. Finally, global gene expression in ES cells as well as iPS cell lines with different developmental competence was analysed as reported²⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 January; accepted 3 July 2009.

Published online 23 July 2009.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
3. Liu, H. *et al.* Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. *Cell Stem Cell* **3**, 587–590 (2008).

4. Li, W. *et al.* Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell* **4**, 16–19 (2009).
5. Liao, J. *et al.* Generation of induced pluripotent stem cell lines from adult rat cells. *Cell Stem Cell* **4**, 11–15 (2009).
6. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
7. Munsie, M. J. *et al.* Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. *Curr. Biol.* **10**, 989–992 (2000).
8. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* **11**, 1553–1558 (2001).
9. Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
10. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
11. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**, 949–953 (2008).
12. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321**, 699–702 (2008).
13. Kim, J. B. *et al.* Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**, 646–650 (2008).
14. Eminli, S., Utikal, J., Arnold, K., Jaenisch, R. & Hochedlinger, K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* **26**, 2467–2474 (2008).
15. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).
16. Wernig, M. *et al.* Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc. Natl Acad. Sci. USA* **105**, 5856–5861 (2008).
17. Shi, Y. *et al.* Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**, 568–574 (2008).
18. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26**, 795–797 (2008).
19. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945–949 (2008).
20. Kaji, K. *et al.* Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* **458**, 771–775 (2009).
21. Woltjen, K. *et al.* *piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).
22. Nagy, A. *et al.* Embryonic stem cells alone are able to support fetal development in the mouse. *Development* **110**, 815–821 (1990).
23. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
24. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25**, 1177–1181 (2007).
25. Takahashi, K., Okita, K., Nakagawa, M. & Yamanaka, S. Induction of pluripotent stem cells from fibroblast cultures. *Nature Protocols* **2**, 3081–3089 (2007).
26. Cheng, J., Dutra, A., Takesono, A., Garrett-Beal, L. & Schwartzberg, P. L. Improved generation of C57BL/6J mouse embryonic stem cells in a defined serum-free media. *Genesis* **39**, 100–104 (2004).
27. Belloch, R., Venere, M., Yen, J. & Ramalho-Santos, M. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell Stem Cell* **1**, 245–247 (2007).
28. Zeng, F. & Schultz, R. M. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Dev. Biol.* **283**, 40–57 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported in part by grants from the China National Basic Research Program 2006CB701500 (to Q.Z.), 2007CB947800 (to F.Z.), 2007CB947700 (to L.W.) and grants from the National Science Foundation of China 30525040 (to Q.Z.) and 30871379/C0607 (to F.Z.).

Author Contributions Q.Z. and F.Z. designed the experiments, supervised lab work, analysed and interpreted data, and wrote the paper; X.Z., W.L., Z.L., L.L., M.T., T.H., J.H., C.G., Q.M. and F.Z. performed experiments; L.L. and W.L. analysed data; W.L. supervised experiments; and X.Z., W.L. and Z.L. contributed to part of the Online Methods section.

Author Information The microarray data in this study have been deposited with the Gene Expression Omnibus repository (www.ncbi.nlm.nih.gov/geo) under accession number GSE16925. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Q.Z. (qzhou@ioz.ac.cn) or F.Z. (fzeng@sjtu.edu.cn).

METHODS

Cell culture. ES cells and iPS cells were cultured as previously described²⁹ on mitomycin-C-treated MEF cells with DMEM (Gibco) plus 15% FBS (Gibco), 1,000 U ml⁻¹ LIF (Chemicon), 2 mM glutamine (Sigma), 1 mM sodium pyruvate (Sigma), 0.1 mM β -mercaptoethanol (Sigma) and 0.1 mM non-essential amino acids. For iPS derivation, DMEM/F12 (1:1, Gibco) and 20% knockout serum (Gibco) was used instead of DMEM with 15% FBS. ES and iPS cells were passaged every 2 days. Oct4-GFP MEF cells (of a B6D2F1 genetic background, that is, F₁ of C57BL/6J \times DBA/2J) used for the generation of iPS cells were isolated from E13.5 embryos containing an Oct4-GFP reporter³⁰, and were cultivated with DMEM plus 10% FBS (Gibco) before retroviral infection. MEF cells for iPS derivation were also made from E13.5 embryos of a C57 \times 129S2 background.

Note that all of the animals that were used for cell culture or for reproductive studies were handled according to the Guidelines for the Care and Use of Laboratory Animals established by the Beijing Association for Laboratory Animal Science.

Retroviral production and infection. Retroviral production and infection followed the previously published protocol²⁵. In brief, the four retroviral vectors (pMXs-Oct4, Sox2, c-Myc and Klf4) were introduced into plat-E cells using lipofectamine 2000 transfection reagent (Invitrogen) according to manufacturer's recommendations. After overnight transduction, the medium was replaced. Another 24 h later, the virus-containing supernatants were collected and filtered through a 0.45- μ m cellulose acetate filter (Millipore), supplemented with 4 μ g ml⁻¹ polybrene (Sigma). Oct4-GFP MEF cells (seeded at 2×10^5 cells per 10-cm culture dish) were incubated with virus-containing supernatants for 48 h with medium change at 24 h before replacement with regular media. Two days after infection, the infected Oct4-GFP MEF cells were replated with 2.5×10^4 cells per 35-mm dish on mitomycin-C-treated MEF feeder layers.

Immunofluorescence analysis and alkaline phosphatase staining. Cells were fixed with 4% paraformaldehyde for 30 min and then permeabilized with 0.5% Triton X-100 for 30 min followed by blocking with 2% BSA (Sigma). Cells were incubated in primary antibody overnight at 4 °C and secondary antibody at room temperature for 1 h. The following antibodies were used: SSEA1 (Chemicon), Oct4 (Santa Cruz), Nanog (Chemicon). Alkaline phosphatase staining was performed with BCIP/NBT Alkaline Phosphatase Colour Development Kit (Beyotime) according to manufacturer's instructions.

Teratoma formation and histological analysis. iPS cells were trypsinized and suspended at 1×10^7 per ml. One-hundred microlitres of the cell suspension was injected into the subcutaneous flanks of the SCID mice. Four to five weeks later, the mice were euthanized and the tumours were fixed and sliced. Sections were stained with haematoxylin and eosin.

Diploid blastocyst injection and tetraploid embryo complementation. Diploid blastocysts were gently flushed out from the uteri of E3.5 timed-pregnant mice with CZB medium³¹. The generation of mice by tetraploid embryo complementation was carried out as previously described²⁹. In brief, two-cell embryos were collected from oviducts of CD-1 females (white coat colour), electrofused to produce one-cell tetraploid embryos that were then cultured in CZB media. Ten to fifteen iPS cells (originally with a B6D2F1 genetic background, or black coat colour) were injected into each tetraploid blastocyst and transferred to CD-1 pseudopregnant recipient females. Embryos derived from tetraploid blastocyst injection (4N) were dissected in handling media on E9.5, E13.5 and the day of birth (E19.5), respectively.

Bisulphite genomic sequencing. Bisulphite treatment of the genomic DNA was performed with the EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's instructions. Oct4 and Nanog promoter regions were amplified with nested primers (Supplementary Table 2). The first round of PCR was performed as follows: 94 °C for 5 min; 35 cycles of 94 °C for 30 s, 59 °C for 45 s, 72 °C for 30 min; and 72 °C for 7 min. The second round of PCR was the same. The PCR products were cloned into pMD18-T vectors (Takara). Ten randomly selected clones were sequenced and analysed.

Karyotype analysis. Karyotype analysis was conducted using standard murine chromosome analysis protocols.

Determination of the SLP by PCR. Sequences for the primer pairs were found on the Mouse Genome Informatics website (<http://www.informatics.jax.org/>). DNA was extracted from tail tips of the mouse or cell pellet in culture and typed as described³². Products were separated by 3% agarose gels and visualized by ethidium bromide staining.

PCR amplification of lineage-specific microsatellite loci. DNA from mouse tails of CD-1, C57, DBA, B6D2F1, 129, one IP14D-1 4N-comp mouse, one offspring of the IP14D-1 4N-comp mouse, one IP14D-1 chimaera, one IP14D-1 chimaera with germline transfer, one IP14D-101 4N-comp mouse, one IP14D-101 chimaera, and from cell lines of Oct4-GFP MEF cells and IP14D-1 were prepared using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's protocol. Amplification of lineage-specific microsatellite DNA was performed according to the procedure described previously³³.

Southern blot analysis. To detect the viral integration of the transgene, genomic DNA was extracted by the classical phenol-chloride extraction method. Twenty-five micrograms of genomic DNA was digested with BglII (Takara) for transgenic c-Myc detection and BamHI (New England Biolabs) for transgene Oct4 detection. Digested DNA was electrophoresed in 0.8% agarose gel overnight and transferred onto positive charged nylon membrane (Millipore). Hybridization was performed at 67 °C with radioactively labelled cDNAs prepared by Prime-a-Gene Labelling System (Promega). Primers used to synthesize the probes were: c-Myc forward, 5'-ACTCCGTACAGCCCTATTTC-3'; c-Myc reverse, 5'-TTCA GCTCGTTCCTCTCT-3'; Oct4 forward, 5'-TGGACACCTGGCTTCAGA-3'; Oct4 reverse, 5'-AGCAGTGACGGGAACAGA-3'.

Reverse PCR. Genomic DNA was extracted from the iPS cell lines IP14D-1, IP14D-6 and IP14D-101, as well as from the tails of IP14D-1 4N-comp, IP14D-6 4N-comp, IP14D-101 4N-comp mice and a normal B6D2F1 mouse using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's protocol. The integration patterns of these samples were identified by reverse PCR (also known as inverse PCR, iPCR) procedures³⁴. Two micrograms of genomic DNA of those mice was digested with two restriction enzymes (BglII and BamHI, New England Biolabs) that have cohesive ends, and cut at least once in pMXs plasmids between the 5'-long terminal repeat (LTR) and the 3'-LTR. The digests were purified by phenol-chloroform extraction and ethanol precipitation, then resuspended in 200 μ l ligation reaction with 1,000 U T4 DNA ligase (New England Biolabs). After incubation at 22 °C overnight, ligated DNA was extracted with phenol-chloroform and precipitated with ethanol, then resuspended in 15 μ l water. The nested iPCR amplifications across the 3'-LTR junction were conducted by using an Advantage 2.0 PCR kit (Clontech). Oct4-specific primers used were: first round: forward, 5'-GCCCCCACTTCACCACACT-3'; reverse, 5'-AGGCCTCGAAGCG ACAGAT-3'; second round: forward, 5'-GAGGTAGACAAGAGAACCTGG AGC-3'; reverse, 5'-GCAAAGTCTCCACGCCAAC-3'.

Reverse transcription and quantitative PCR analysis. To test the expression of pluripotent genes with endogenous and transgenic origin, total RNA was isolated using TRIzol reagent (Invitrogen) and first-strand complementary DNA was synthesized using SuperScript III Reverse Transcriptase (Invitrogen) and oligo-dT (Promega) according to manufacturer's instructions. qPCR was performed using SYBR green (Sigma) on an ABI 7000 instrument. Primer sequences for each gene are listed in Supplementary Table 3.

Microarray analysis. Microarray analysis was performed as previously reported^{28,35}. Total RNA was extracted from three replicates of each cell line, including the following: CL11 (a B6D2F1 background ES cell line from which we recently succeeded in generating live 4N-comp animals, provided by S. Gao, National Institute of Biological Sciences), IP14D-1, IP14D-101 (4N-comp capable iPS cell lines), IP20D-3 (2N germline transmittable) and MEF (Oct4-GFP) cells using TRIzol reagent (Invitrogen). RNA mass and size distribution were determined using the Agilent Bioanalyser with RNA 6000 Nano LabChips, and 2.5 μ g of total RNA was used as initial template for biotin-labelled cRNA amplification with GeneChip One-Cycle labelling kit (Affymetrix). Fifteen micrograms per replicate of fragmented cRNA was hybridized to MOE430 2.0 GeneChips (Affymetrix) then washed and stained on fluidics stations and scanned using a GCS3000 scanner according to the manufacturer's instructions. Expression Console (with MAS5) was used to quantify microarray signals with default analysis parameters and global scaling to target mean = 150. Scatter-plot was used to display the expression values for all probe sets from genome-wide transcription profiling derived from the different samples. Hierarchical clustering was performed using the NIA Array Analysis Tool (<http://lgsun.grc.nia.nih.gov/ANOVA/index.html>) to visualize inter-relationships between different cell lines using all microarray data and the subset of significantly different genes by multi-class analysis of variance (ANOVA) with multiple testing correction. The expression profiles of selected key pluripotent marker genes were also plotted for comparison. The microarray data for each sample were deposited with the Gene Expression Omnibus repository (www.ncbi.nlm.nih.gov/geo/) under accession number GSE16925.

29. Nagy, T., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2003).
30. Yoshimizu, T. et al. Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* **41**, 675–684 (1999).
31. Chatot, C. L., Ziomek, C. A., Bavister, B. D., Lewis, J. L. & Torres, I. An improved culture medium supports development of random-bred 1-cell mouse embryos *in vitro*. *J. Reprod. Fertil.* **86**, 679–888 (1989).
32. Zhou, Q., Juneau, A., Brochard, V., Adenot, P. & Renard, J. P. Developmental potential of mouse embryos reconstructed from metaphase embryonic stem cell nuclei. *Biol. Reprod.* **65**, 412–419 (2001).
33. Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res.* **18**, 4123–4130 (1990).

34. Olivares, E. C. *et al.* Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nature Biotechnol.* **20**, 1124–1128 (2002).
35. Zeng, F., Baldwin, D. A. & Schultz, R. M. Transcript profiling during preimplantation mouse development. *Dev. Biol.* **272**, 483–496 (2004).

Adult mice generated from induced pluripotent stem cells

Michael J. Boland^{1*}, Jennifer L. Hazen^{1*}, Kristopher L. Nazor^{1*}, Alberto R. Rodriguez², Wesley Gifford³, Greg Martin², Sergey Kupriyanov² & Kristin K. Baldwin¹

Recent landmark experiments have shown that transient overexpression of a small number of transcription factors can reprogram differentiated cells into induced pluripotent stem (iPS) cells that resemble embryonic stem (ES) cells^{1–7}. These iPS cells hold great promise for medicine because they have the potential to generate patient-specific cell types for cell replacement therapy and produce *in vitro* models of disease, without requiring embryonic tissues or oocytes^{8–10}. Although current iPS cell lines resemble ES cells, they have not passed the most stringent test of pluripotency by generating full-term or adult mice in tetraploid complementation assays^{3,11}, raising questions as to whether they are sufficiently potent to generate all of the cell types in an organism. Whether this difference between iPS and ES cells reflects intrinsic limitations of direct reprogramming is not known. Here we report fertile adult mice derived entirely from iPS cells that we generated by inducible genetic reprogramming of mouse embryonic fibroblasts. Producing adult mice derived entirely from a reprogrammed fibroblast shows that all features of a differentiated cell can be restored to an embryonic level of pluripotency without exposure to unknown ooplasmic factors. Comparing these fully pluripotent iPS cell lines to less developmentally potent lines may reveal molecular markers of different pluripotent states. Furthermore, mice derived entirely from iPS cells will provide a new resource to assess the functional and genomic stability of cells and tissues derived from iPS cells, which is important to validate their utility in cell replacement therapy and research applications.

Historically, the only way to generate an adult mammal was by fertilization. Advances in somatic cell nuclear transfer (SCNT) have now produced genetically identical mouse clones from a variety of differentiated cell types, from fibroblasts to neurons^{12,13}. Similarly, genetically identical adult mice may be derived from ES (or SCNT-ES) cells by tetraploid blastocyst complementation, in which all adult tissues derive from the stem cell line whereas extraembryonic tissues are supplied by the tetraploid cells^{14,15}. For unknown reasons, current iPS cell lines have not generated adult or full-term mice in tetraploid complementation assays. This finding, and recent reports of reproducible gene expression differences between iPS and ES cells, suggests that direct reprogramming may be insufficient to restore differentiated cells to full pluripotency, as measured by ES cell equivalence^{3,11,16}. Autonomous generation of mice from iPS cells would validate direct reprogramming as equivalent to reprogramming by SCNT, establish iPS cells as functional substitutes for ES cells, and provide a new method to generate adult mice from differentiated cells.

To conclusively demonstrate that iPS cell lines can generate adult mice in tetraploid complementation assays, we designed a genetic

marking strategy to distinguish between host blastocyst and iPS-derived cells. We established mouse embryonic fibroblasts (MEFs) from animals generated by a cross of two mouse lines (*Pcdh21*/Cre and Z/EG, Fig. 1a). The Z/EG transgene labels most cells in an animal with a visible marker (β -geo, a fusion of the β -galactosidase and neomycin genes)¹⁷, whereas the *Pcdh21*/Cre modification results in Cre expression in rare neuronal subtypes, but not in ES cells¹⁸. Cre expression causes excision of the floxed β -geo gene, resulting in green fluorescent protein (GFP) expression in olfactory bulb mitral cells, a feature we exploit later (Fig. 1a).

We reasoned that the inappropriate expression of reprogramming genes during development could inhibit embryonic and postnatal development. Therefore, we designed a drug-inducible lentiviral reprogramming strategy to achieve tight control of transgene expression in iPS cells and their derivatives (Fig. 1b)¹⁹. The four original reprogramming factors (*Oct4* (also known as *Pou5f1*), *Sox2*, *Klf4* and *c-Myc*) were placed under control of the tetO promoter, which is activated by the reverse tetracycline transactivator (rtTA) protein in the presence of the tetracycline analogue doxycycline (dox). We used an enhanced version of the rtTA transcriptional activator protein (rtTAM2.2) that induces higher gene expression levels than the rtTAM2 protein²⁰. To promote complete reprogramming and to facilitate isolation of fully reprogrammed iPS cells we exposed MEFs to the histone deacetylase inhibitor valproic acid (VPA), which has been reported to enhance reprogramming efficiency and to select against incompletely reprogrammed cells by inhibiting cell division^{21,22} (see Methods).

Reprogramming of *Pcdh21*/Cre–Z/EG fibroblasts by this method resulted in ES-cell-like colonies after five (with dox plus VPA) to seven (dox only) days of dox induction (Fig. 1c, d and Methods). No colonies emerged in the absence of dox, which demonstrates both the inducibility and specificity of our system. We subcloned 21 colonies from the dox-plus-VPA-treated cells, which we refer to as iMZ iPS cell lines.

At present, there is no established method to select iPS cells that will contribute to all of the tissues of an organism. To prioritize the iMZ cell lines for tetraploid complementation assays, we assessed lines for similarity to ES cells by morphology, proliferation rate, expression of pluripotency markers and ability to generate embryoid bodies (Figs 1d, 2d and Supplementary Figs 2 and 3). We also exploited our cell-type-specific Cre line to determine whether embryoid bodies made from our iMZ lines could generate cells that resembled olfactory bulb mitral cells on the basis of neuronal morphology and GFP expression (Fig. 1e and Supplementary Fig. 3). Using these criteria, we selected 12 candidate lines for karyotype analysis (Supplementary Figs 4, 5 and Methods). To establish the pluripotency of lines

¹Department of Cell Biology, ²Mouse Genetics Core Facility, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ³Medical Scientist Training Program, University of California, San Diego, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

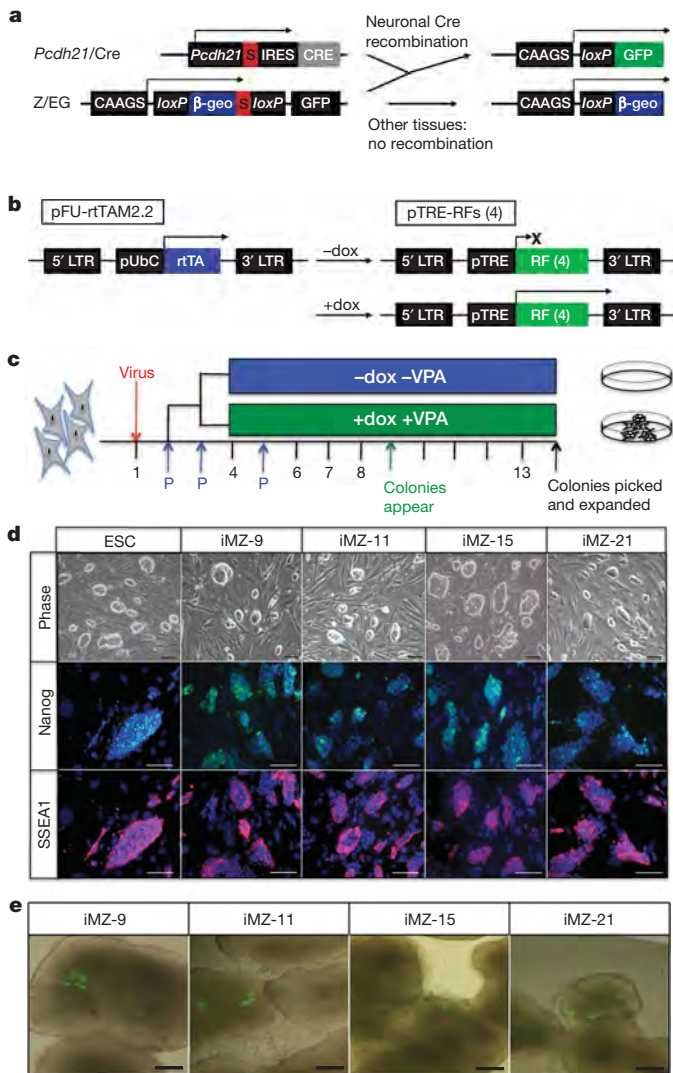


Figure 1 | Generation of iPS cells. **a**, Genetic marking strategy. Z/EG cells express β -geo. Rare neurons express *Pcdh21*/Cre causing GFP expression. **b**, Drug inducible reprogramming. Reprogramming factors (RFs; *Oct4*, *Sox2*, *c-Myc* or *Klf4*) are controlled by the dox-inducible promoter (pTRE). The rtTAM2.2 is constitutive (pUbc; ubiquitin promoter). **c**, Reprogramming timeline (x-axis: days after transduction). Dox and VPA treatment began on day 4 and ended on day 23. P, passage. **d**, Brightfield images (top row) of iMZ cell lines and ES cells (ESC). Immunofluorescence staining (middle and bottom rows) of iMZ lines for pluripotency markers Nanog (green), SSEA1 (red) and nuclei (stained with 4,6-diamidino-2-phenylindole (DAPI), blue). **e**, iMZ-9 cell-derived embryoid bodies contain GFP-positive cells. Scale bars, 100 μ m.

generated in this model, we tested three euploid lines (iMZ-9, -11 and -21) in diploid blastocyst injection assays. All tested lines contributed to chimaeric mice on the basis of coat colour. Line iMZ-9 iPS cells generated several mice with nearly 100% agouti fur (Supplementary Fig. 6). All iPS cell lines contributed to all germ layers based on expression of the β -geo transgene in several tissues (Supplementary Fig. 7), production of GFP⁺ cells in the olfactory bulb (Fig. 2c), and the germline transmission of lines iMZ-9 and iMZ-11 (data not shown).

Southern blot analyses showed that lines iMZ-9 and iMZ-21 have identical patterns of proviral insertions and thus, these lines are probably derived from the same initial transduced MEF (Fig. 3b and Supplementary Fig. 9). After induction of transgene expression during iPS cell derivation, iMZ MEFs were split only once. Therefore, these two independently isolated iPS cell lines potentially experienced different stochastic events during reprogramming, which can confer different

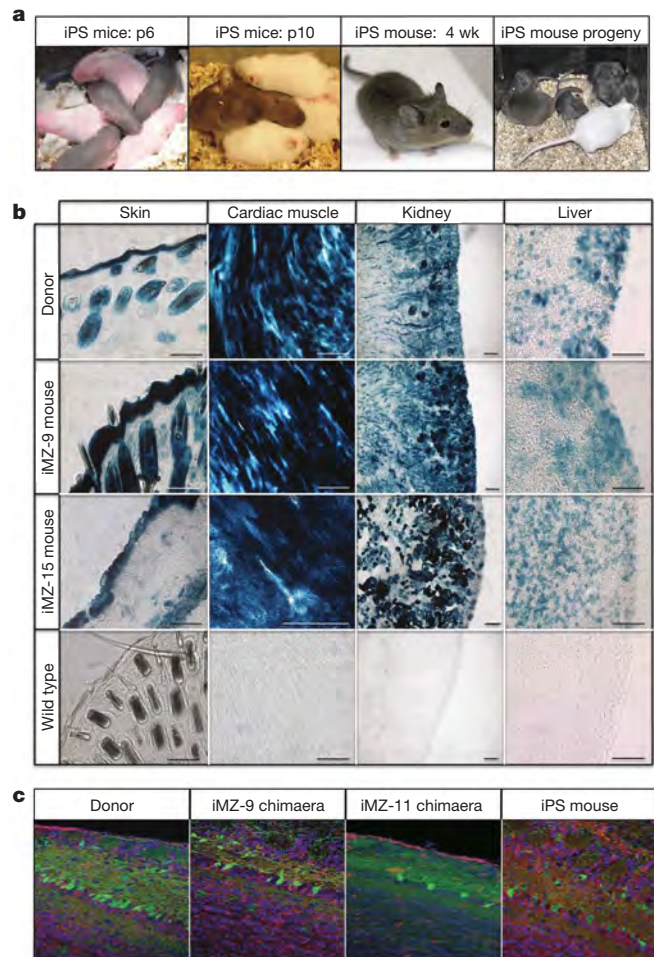


Figure 2 | Adult mice derived from iMZ cell lines. **a**, Far left, iPS mice (postnatal day (p) 6) exhibit pigmented skin in comparison to CD-1 mice. Middle left, iPS mice (postnatal day 10) and age-matched albino CD-1 pups. Middle right, adult iMZ-9 iPS mouse (4 weeks) is morphologically normal. Far right, germline transmission by 12-week iMZ-9 mouse (left) shown with 2-week-old progeny (agouti) and mother (white). **b**, Tissue sections from a postnatal day 10 *Pcdh21*/Cre-Z/EG mouse (top), iPS mice derived from iMZ-9 and iMZ-15 cell lines (middle) and wild-type mouse (bottom) stained with X-gal (blue). Scale bars, 100 μ m. **c**, Contribution to the olfactory bulbs of postnatal day 0 chimaeric mice (iMZ-9 and iMZ-11) and an iMZ-9 iPS mouse. Sections were stained for GFP (green, mitral cells), β -galactosidase (red, Z/EG cells) and nuclei (blue, TOTO-3). Original magnification, $\times 400$.

epigenetic alterations and variable developmental potential on otherwise identical cell lines^{23,24}. For these reasons, we refer to iMZ-9 and iMZ-21 as independent lines.

To determine whether these iMZ lines could generate full-term or adult mice, we performed tetraploid complementation assays. In a series of independent experiments, we injected albino tetraploid blastocysts with cells derived from iMZ-9, iMZ-11, iMZ-15 and iMZ-21 and from two other iPS cell lines (iNZ-3 and iNZ-19) (Table 1 and Supplementary Methods). We performed Caesarean section on the evening before scheduled delivery and obtained breathing pups, termed iPS mice, with normal morphology from lines iMZ-9 (four viable pups, three either non-viable or cannibalized after fostering, 13 apparently viable on embryonic day (E) 16.5 or E17.5), iMZ-21 (ten viable pups, eight non-viable), iMZ-15 (one live pup with a herniated umbilical cord, one full-term pup with respiratory failure, one cannibalized) and iMZ-11 (one live pup, later cannibalized). Lines iNZ-3 or iNZ-19 did not generate full-term pups (Table 1 and Fig. 2a).

Most surviving pups show no obvious morphological abnormalities (Fig. 2a). Non-viable pups typically presented with difficulty breathing,

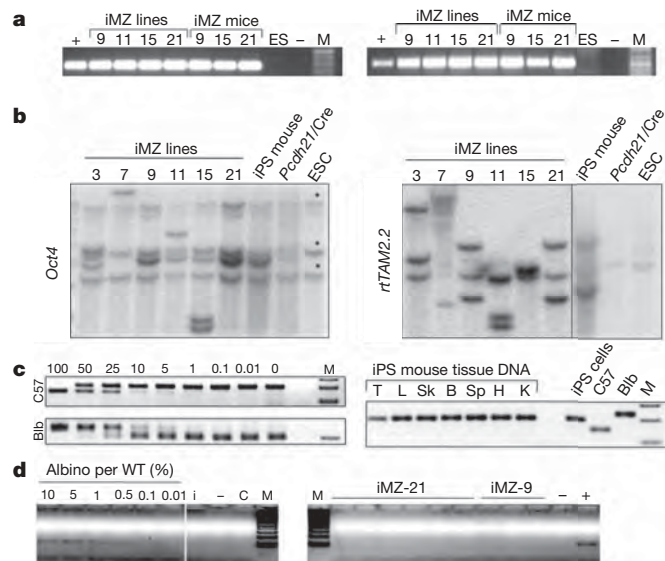


Figure 3 | Genetic analyses of iPS mice. **a**, PCR assay of genomic DNA for the Cre (left) and Z/EG (right) genetic insertions in iMZ iPS cells and mice. The positive control (+) is *Pcdh21*/Cre–Z/EG tail DNA. No bands are seen in the negative control (–) (no DNA) and ES cell lanes. M denotes molecular mass marker. **b**, Southern blots of genomic DNA from iMZ cell lines and an iMZ-9 iPS mouse show similar patterns of insertions. Probes were coding sequences of the *Oct4* (left) or *rtTAM2.2* (right) genes. *Pcdh21*/Cre mice and wild-type ES cells are controls. Asterisks denote endogenous bands. **c**, Microsatellite PCR assay for tetraploid cells. Band size distinguishes iMZ cells from tetraploid host strains (C57BL/6J-*Tyr^{c-2J}* (C57) and BALB/cByJ (Bib)). Left, DNA titration curve demonstrates 5% detection limit. Right, analysis of DNA from tissues derived from an iPS mouse (iMZ-9). B, brain; H, heart; K, kidney; L, liver; Sk, skin; Sp, spleen; T, thymus. **d**, Albino allele PCR assay. Left, albino *Tyr^{c-2J}* allele PCR assay can detect 0.5% tetraploid cell DNA diluted into iMZ DNA. C, CD-1; i, iMZ-9; –, H₂O control. The C57BL/6J-*Tyr^{c-2J}* allele is expected in 75% of tetraploid blastocysts. Right, tail DNA from different pups derived from iMZ iPS lines as noted. –, H₂O control; +, C57BL/6J-*Tyr^{c-2J}* DNA; WT, wild type.

which is common in tetraploid complementation experiments performed with ES cells¹⁴. The efficiency of generating iPS mice ranged from 0.3% to 13%, which is similar to published efficiencies for ES and SCNT-ES cells (Table 1)^{15,25,26}. Notably, iMZ-9 iPS cells reproducibly generated adult mice in several experiments.

Although tetraploid cells rarely contribute to the embryo proper beyond mid-gestation^{25,27}, to conclusively demonstrate that these iPS mice derived entirely from the iMZ cell lines we analysed coat and eye colour and performed histological staining. The coat and eye colour of the iPS mice (agouti, pigmented) differs from both the albino tetraploid host blastocyst and the albino recipient female. As expected, all pups exhibited pigmented eyes and uniformly agouti fur (Fig. 2a), in contrast to the variation in coat colour observed in the diploid chimaera assays (Supplementary Fig. 6). Similarly, cells derived from iMZ lines stain positive for β -galactosidase, whereas blastocyst-derived cells do not. Intact tissues and tissue sections representing all three germ layers showed positive staining for

β -gal (Fig. 2b and Supplementary Fig. 7). We observed no histological differences between iPS mice and the *Pcdh21*/Cre–Z/EG mouse strain, whereas staining was clearly different from iMZ chimaeric animals (Fig. 2b, c and Supplementary Fig. 7). Furthermore, immunofluorescence analyses of the olfactory bulbs of iPS mice showed that all cells express β -gal (except for the mitral cells which express GFP, as expected) (Fig. 2c).

To exclude minor contributions of tetraploid host blastocyst cells during iPS mouse development, we designed sensitive PCR assays to detect microsatellite markers and albino mutations of the *Tyr* gene that differ between host blastocysts and iMZ cells (Fig. 3c, d). Microsatellite analyses of several tissues from an iMZ-9 mouse showed no contribution from host blastocyst cells in the microsatellite assay (Fig. 3c). Similar results were observed with tail tissue from an iMZ-15 mouse (Supplementary Fig. 8). We also performed tests for the albino mutation on DNA from nine individual mice derived from lines iMZ-9 and iMZ-21 and detected no contribution of host blastocyst cells (Fig. 3d). These data strongly support the conclusion that the iMZ lines are capable of generating all cell types of adult mice in tetraploid complementation assays, at least to the level that is typical of ES cells.

Notably, genetic analyses rule out contamination of our iPS cell lines by pre-existing ES cells. PCR experiments demonstrated that all iMZ lines and iPS mice carry the *Pcdh21*/Cre and Z/EG genetic modifications, which do not co-exist in any known ES cell lines (Fig. 3a). Similarly, Southern blot analyses with probes for *Oct4* and *rtTAM2.2* show that the patterns of proviral integration of the iPS mice are identical to the patterns of the iPS cell lines from which they derived. Furthermore, the relative intensity of the *Oct4* proviral insertions and endogenous *Oct4* bands is similar as expected if iPS mice derive predominantly from iPS cells (Fig. 3b).

A final test of pluripotency is to establish the germline contribution and production of viable offspring. In crosses with albino female mice, male iPS mice derived from lines iMZ-9 and iMZ-21 exhibit germline transmission as evidenced by the production of 100% agouti pups and expression of the Z/EG allele in the expected number of progeny (Fig. 2a and Supplementary Fig. 8). Taken together, these data demonstrate that direct reprogramming of MEFs with four factors can generate iPS cell lines that possess full pluripotency as measured by the production of fertile adult mice in tetraploid complementation assays.

Several features of our experiments could be responsible for the enhanced pluripotency of our iPS cell lines. First, the high levels of transgene induction afforded by rtTAM2.2 and the extended duration of reprogramming factor expression in our protocol may more completely reprogram cells. Second, transgene expression in our iPS cell lines is regulated by a dox-inducible promoter, which may help to prevent inappropriate expression of reprogramming factors during embryonic development. In support of this idea, quantitative PCR with reverse transcription (qRT-PCR) experiments demonstrate that proviral transgenes are nearly completely silent in iPS cells in the absence of dox (Supplementary Fig. 11). Furthermore, the lines that generate iPS mice most robustly (iMZ-9 and iMZ-21) have reduced expression of all four reprogramming factors, whereas the less efficient lines (iMZ-11 and iMZ-15) have detectable expression of *Klf4*

Table 1 | Summary of blastocyst injections

| Cell line | Chimaera (2n) | Blastocysts injected (4n) | Day fetuses removed | Live fetuses (% inj.) | Live pups (% inj.) | Suckling pups (% inj.) |
|-----------|---------------|---------------------------|---------------------|-----------------------|--------------------|------------------------|
| iMZ-9 | Y | 172 | E16.5 | 11 (6%) | – | – |
| iMZ-9 | Y | 25 | E17.5 | 2 (8%) | – | – |
| iMZ-9 | Y | 195 | E18.5 | – | 7 (4%) | 4 (2%) |
| iMZ-21 | Y | 140 | E18.5 | – | 18 (13%) | 10 (7%) |
| iMZ-15 | n.d. | 257 | E18.5 | – | 3 (1%) | 0 |
| iMZ-11 | Y | 338 | E18.5 | – | 1 (0.3%) | 0 |
| iNZ-3 | n.d. | 120 | E18.5 | – | 0 | 0 |
| iNZ-19 | Y | 131 | E18.5 | – | 0 | 0 |

Y denotes iPS cell lines that contributed to diploid chimaeras; n.d. denotes lines that were not tested. Recipients were either euthanized at E16.5 or E17.5 or delivered by Caesarean at E18.5. The numbers of live fetuses, live pups and suckling pups surviving after postnatal day 1 are shown with the percentage of total blastocysts (% injection (inj.)).

and/or *Oct4* in the absence of dox. Third, previously reported reprogramming experiments either did not use VPA or used short VPA treatments but did not report tetraploid complementation experiments. Prolonged VPA treatment in our experimental protocol may have allowed resetting of the epigenome to a chromatin state more similar to that of ES cells. The timing and extent of passaging, the genetic background of reprogrammed MEFs and our iPS cell selection criteria may have contributed to the enhanced pluripotency of our lines.

We cannot exclude the possibility that the enhanced pluripotency of the iPS cell lines we report here is a result of reprogramming of a rare cell type or a particular pattern of proviral insertion. However, we have established three iPS cell lines with distinct proviral integration patterns that can generate full-term mice, suggesting that several individual MEF cells can lead to fully pluripotent iPS cells and that a particular insertion pattern is not required. At present, the two iMZ lines that generate fertile adult mice derive from a single infected cell. Until other adult mice are generated we cannot exclude models requiring a rare cell type or particular proviral insertion.

The generation of fertile adult mice from iPS cells acts as an important proof-of-principle validation that reprogramming technology can produce iPS cells with functional equivalence to ES cells in developmental assays. These data demonstrate that direct reprogramming with four factors can recapitulate the reprogramming capacity of the oocyte and show that non-genomic components of differentiated cells, such as mitochondria, do not impede reprogramming to full pluripotency. We speculate that comparing iPS cell lines that generate iPS mice with those that cannot generate mice but satisfy other criteria of pluripotency (for example, chimaerism and germline contribution) may reveal important molecular differences associated with states of pluripotency.

METHODS SUMMARY

The reported iPS cells were derived from E13.5 MEFs using dox-inducible lentiviruses encoding *Oct4*, *Sox2*, *Klf4* and *c-Myc* as previously described²⁸, except that induction was mediated by *rtTAM2.2* and VPA was included. Reprogrammed iPS cell lines were characterized by immunofluorescence (for SSEA1, Nanog, *Oct4* and *Sox2*). *In vitro* differentiation into embryoid bodies by aggregation in suspension and treatment with *all-trans* retinoic acid was used to assess the ability of iMZ lines to generate rare GFP⁺ cells. Karyotype was examined by analysis of metaphase spreads prepared by the hanging-drop method. Chimaeric mice were produced by injection of euploid iPS cell lines into diploid blastocysts²⁹. iPS mice were produced by the injection of iPS cells into tetraploid blastocysts generated by electrofusion of two-cell embryos according to established methods^{15,29}. Intact tissues and tissue sections from chimaeras or iPS mice were stained with X-gal or antibodies for β -galactosidase and GFP. Genetic analyses of iPS mice were performed by Southern blotting for proviral insertions, genotype analysis for the *Pcdh21/Cre* and Z/EG alleles, and PCR for microsatellite markers and the *Tyr⁺* albino mutation using standard methods. Residual expression of virally encoded transgenes was examined by qRT-PCR using lentiviral-specific primers.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 June; accepted 23 July 2009.

Published online 2 August 2009.

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
3. Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
4. Park, I. H., Lerou, P. H., Zhao, R., Huo, H. & Daley, G. Q. Generation of human-induced pluripotent stem cells. *Nature Protocols* **3**, 1180–1186 (2008).
5. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
6. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).

7. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4**, 381–384 (2009).
8. Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277–280 (2009).
9. Park, I. H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
10. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).
11. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
12. Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).
13. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
14. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
15. Eggan, K. & Jaenisch, R. in *Mammalian and Avian Transgenesis—New Approaches* (eds Pease, S. & Lois, C.) Ch. 3 41–68 (Springer, 2006).
16. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
17. Novak, A., Gup, C., Yang, W., Nagy, A. & Lobe, C. G. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon cre-mediated excision. *Genesis* **28**, 147–155 (2000).
18. Nagai, Y., Sano, H. & Yokoi, M. Transgenic expression of Cre recombinase in mitral/tufted cells of the olfactory bulb. *Genesis* **43**, 12–16 (2005).
19. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
20. Go, W. Y. & Ho, S. N. Optimization and direct comparison of the dimerizer and reverse tet transcriptional control systems. *J. Gene Med.* **4**, 258–270 (2002).
21. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. *Nature Methods* **6**, 363–369 (2009).
22. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26**, 795–797 (2008).
23. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52 (2009).
24. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
25. Eakin, G. S., Hadjantonakis, A. K., Papaioannou, V. E. & Behringer, R. R. Developmental potential and behavior of tetraploid cells in the mouse embryo. *Dev. Biol.* **288**, 150–159 (2005).
26. Eggan, K. & Jaenisch, R. Differentiation of F1 embryonic stem cells into viable male and female mice by tetraploid embryo complementation. *Methods Enzymol.* **365**, 25–39 (2003).
27. Mackay, G. E. & West, J. D. Fate of tetraploid cells in 4n \leftrightarrow 2n chimeric mouse blastocysts. *Mech. Dev.* **122**, 1266–1281 (2005).
28. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26**, 916–924 (2008).
29. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* 453–506 (Cold Spring Harbor Laboratory Press, 2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We wish to thank S. Carlson for help with animal husbandry, K. Spencer for help with imaging, and D. Trajkovic for help with histology. We thank G. Joyce, U. Mueller, L. Stowers, A. Patapoutian and A. Maximov for critical reading of the manuscript. We thank M. Mayford for the gift of *rtTAM2.2*, A. Maximov for the lentiviral vector backbones, and members of A. Kralli's laboratory for assistance with qPCR. We thank R. Axel for supporting the generation of the *Pcdh21/Cre* mouse strain, which was a gift. This work was supported by a Pew Scholars Award (K.K.B.) and grants from the California Institute of Regenerative Medicine, the Whitehall Foundation, the O'Keefe Foundation, and the Shapiro Family Foundation.

Author Contributions M.J.B., J.L.H. and K.L.N. designed and performed experiments, analysed data and edited the manuscript. A.R.R. performed blastocyst injections, Caesarean sections and cross fostering. W.G. assisted in designing and generating lentiviral constructs. G.M. assisted with cell culture. S.K. designed and performed experiments, analysed data and edited the manuscript. K.K.B. conceived of the experimental design, performed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.K.B. (kbaldwin@scripps.edu) or K.L.N. (kitnazor@scripps.edu).

METHODS

Generation of *Pcdh21*/Cre-Z/EG and *Nes*/Cre-Z/EG mice. To generate the *Pcdh21*-Cre mice, an internal ribosome entry site (IRES)-Cre recombinase together with an FRT-flanked *neo* expression cassette (FRT-*neo*-FRT) was inserted into the *Pcdh21* locus immediately after the translational stop sequence by homologous recombination in ES cells. ES cell colonies were screened and confirmed by Southern blot. Positive colonies were used to generate chimaeric mice and these mice or their *Pcdh21*/Cre positive offspring were crossed to Z/EG mouse lines to generate the *Pcdh21*/Cre-Z/EG mouse strain. No ES cells containing both modifications have been produced. Mice retain the FRT-*neo*-FRT cassette. Mouse genotypes were confirmed by PCR for the wild-type *Pcdh21* allele, the *Pcdh21*-Cre knock-in allele, and β -geo. Primer sequences and PCR conditions are available on request. The *Nes* (also known as *Neurod6* or *Math2*)/Cre mouse line labels post-mitotic neurons in various brain regions^{30,31}. We crossed this line to the Z/EG line to produce *Nes*/Cre-Z/EG mice from which the control iNZ fibroblasts were derived.

Generation of lentiviral constructs. All lentiviral shuttle vectors were generated from a modified version of the FUGW vector^{19,32}. To generate pFU-rtTA, the *rtTAM2.2* gene²⁰ was cloned from the pWG020 vector into the XbaI and BamHI sites of the viral vector MCS. A linker containing an extra BamHI site, a kozak sequence and a start codon was inserted into the XbaI site. The dox-inducible lentiviral construct, pFT-MCS, was generated by replacing the human ubiquitin C promoter with seven tetO repeats followed by a minimal CMV promoter. The dox-inducible promoter was amplified from pTRE-d2eGFP (BD Biosciences, Clontech) and cloned into the PacI and XbaI sites of the FUGW-derived vector. The coding sequences of *Oct4*, *Sox2*, *c-Myc* and *Klf4* were ligated into pFT-MCS. *Oct4*, *Sox2* and *Klf4* were inserted into the EcoRI site. *c-Myc* was inserted using the XbaI and BamHI sites.

Production of lentivirus. Virus was produced in HEK293T cells by calcium phosphate co-transfection of lentiviral shuttle vectors with the pCMVΔ8.9 and pVSVg viral packaging vectors. Virus was collected at 24, 48 and 72 h post-transfection and concentrated by ultracentrifugation (2 h at 112,000g) at 4 °C.

Generation of iPS cells. MEFs were prepared from *Pcdh21*/Cre-Z/EG (iMZ lines) or *Nes*/Cre-Z/EG (iNZ lines) E13.5 embryos. For the generation of iMZ cell lines, after 24 h in culture, individual wells of ~300,000 MEFs were transduced with lentiviruses (day 1) and split 1:2 (day 2) and 1:3 (day 3) to generate six wells of transduced MEFs. On post-transduction day 4, dox (10 μ g ml⁻¹) was added to four wells to induce expression of reprogramming genes; three of these wells were also treated with VPA (1.9 mM). The remaining two wells were treated with nothing, or VPA alone. To maintain conditions for optimal MEF growth and viability, on post-transduction day 5, the three dox and VPA wells were expanded into a 15 cm² dish whereas the other three conditions were expanded to 10 cm² dishes. On post-transduction day 9 (5 days after dox induction), colonies were observed in the dox plus VPA wells; colonies emerged in the dox only wells on day 11 (7 days after dox induction). No colonies emerged in the absence of dox, with or without VPA. On post-transduction day 12, eight small colonies (iMZ-1–8) were isolated from the dox-plus-VPA plate by aspiration into a P10 (10 μ l) pipette tip, brief trypsinization, and plating to mitotically inactive MEF feeders in a 96-well dish. On day 14, further colonies were isolated, of which 12 were expanded (lines iMZ-9–21) and characterized further. Lines iMZ-1, -4, -5, -6, -8 and -16 did not grow or proliferated more slowly than ES cell controls, so these were not maintained. On post-transduction day 17, all putative iPS cell lines were transferred to fresh feeders in single 96-well plates. Cell lines were expanded into two wells on post-transduction day 21. Cells were maintained in dox plus VPA until post-transduction day 23, when both were removed. All cell lines appeared to maintain ES-cell-like morphology and proliferation rates. Lines were expanded to 24-well plates on day 24. Subsequently, cell lines were banked and maintained as described in the cell culture conditions section.

For the generation of iNZ lines, initial reprogramming conditions were identical to those of iMZ lines except that cells were split 1:2 on post-transduction day 3 to generate four wells (dox plus VPA, dox only, VPA only and no treatment). Cells were split once after dox and VPA addition. In this experiment, colonies were observed on post-transduction day 11 in wells with dox plus VPA and dox only, but not in wells lacking dox, as expected. VPA treatment was halted for 1 day at day 12 to allow proliferation and then resumed, whereas dox treatment was continuous. On day 14, 39 colonies were observed in dox plus VPA (20) and dox only (19) wells. This would represent a reprogramming efficiency of 0.03% (20 colonies per 75,000 initial fibroblasts), which does not take into account the transduction efficiency for all five viruses, or the expansion of clones with identical insertions. After splitting the dox-plus-VPA wells, 32 colonies from the dox-plus-VPA wells were isolated and 19 maintained proliferation at rates similar to ES cells. These lines are called iNZ-1–19.

Cell culture conditions. ES and iPS cells were maintained on mitotically inactivated MEF feeders in 85% DMEM, 15% ES cell qualified FBS (Gibco), 1 mM L-glutamine, 0.1 mM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 1,000 U of ESGRO per ml (Chemicon), 100 U ml⁻¹ penicillin and 10 μ g ml⁻¹ streptomycin. MEF feeders were maintained on 0.1% gelatin-coated dishes in 70% DMEM, 20% medium 199, 10% FBS and 100 U ml⁻¹ penicillin/streptomycin. All cells were kept at 37 °C in a humidified environment at 5% CO₂. Embryoid bodies were aggregated in suspension using ultra-low attachment surfaces (Corning) in ES cell medium lacking ESGRO and 2-mercaptoethanol and treated with 2 \times 10⁻⁶ M *all-trans* retinoic acid (Sigma) from days 4–10.

Southern blotting. Genomic DNA was prepared using the DNAeasy Blood and Tissue Kit (Qiagen). Eight micrograms of DNA were digested with PvuII (*Oct4*), BamHI (*Sox2*, *Klf4* and *c-Myc*) or EcoRI (*rtTAM2.2*), resolved on 0.8% agarose gels, transferred to Hybond-N+ membrane (Amersham Biosciences) and hybridized with radiolabelled-probe at 65 °C. Probes were generated using the Prime-It II Random Primer Labelling Kit (Stratagene). Images were captured on a Typhoon 8600 Variable Mode Imager and analysed with ImageQuant 5.2 software. The open reading frame (ORF) of each gene served as template for probe synthesis. *Oct4* ORF (NCBI accession NM_013633) = 1,058 base pairs (bp); *Sox2* ORF (NM_011443) = 959 bp; *c-Myc* ORF (NM_010849) = 1,364 bp; *Klf4* ORF (NM_010637) = 1,451 bp; and *rtTAM2.2* ORF = 1,490 bp. After hybridization, blots were successively washed with 2 \times SSC/0.1% SDS at room temperature and 0.2 \times SSC/0.1% SDS at 65 °C.

Immunofluorescence analyses of cell lines. Cells were fixed with 4% paraformaldehyde (PFA) at room temperature for 20 min, blocked and permeabilized for 1 h at room temperature in PBS/Triton X-100 (0.1%), incubated overnight at 4 °C in primary antibodies against Oct4 (Santa Cruz Biotechnology, 1:100), SSEA1 (Developmental Studies Hybridoma Bank, 1:500), Nanog (Cosmo Bio Co., 1:50), Sox2 (R&D Systems, 1:50), washed in blocking solution for 3 \times 15 min, incubated for 30 min at room temperature with fluorescence-conjugated secondary antibodies (Alexa). Nuclei were labelled with DAPI or TOTO-3 (Molecular Probes, Invitrogen). Images were collected on an Olympus BX60M microscope and analysed with MetaMorph software.

Analyses of tissues and tissue sections. Whole tissues were dissected and placed directly into X-gal staining buffer (100 mM sodium phosphate, pH 7.3, 2 mM MgCl₂, 0.01% sodium deoxycholate, 0.02% NP-40, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 1 mg ml⁻¹ X-gal) and incubated at 37 °C until staining of controls was evident (several hours). Tissues were washed in PBS and preserved in 4% PFA/PBS fixative. For sections, tissues were collected and fixed with 4% PFA/PBS for 1 h at 4 °C, 30% sucrose protected overnight at 4 °C, OCT-embedded and cut into 30- μ m sections using a Leica CM 3050S Cryostat. Sections were air-dried on charged slides for 20 min and fixed in 4% PFA for 7 min. Sections were then X-gal-stained for 2–3 h at 37 °C, mounted and imaged on an Olympus AX70 microscope, and analysed with Spot imaging software. Alternatively, brain slices were co-stained with primary antibodies against LacZ (Promega, 1:500) and GFP (Invitrogen, 1:500) and imaged on an Olympus Fluoview FV500 LSM microscope. Images were analysed using MetaMorph software.

Microsatellite PCR assay. The length of the microsatellite detected by the *D12Mit136* primer pair is different in each of the *Pcdh21*/Cre-Z/EG, C57BL/6J-*Tyr^{c-2j}* and BALB/cByJ mouse strains. The genotype of the host tetraploid blastocysts varied in experiments but in each case tetraploid blastocysts will carry either the BALB/c allele or both the BALB/c and the C57BL/6J alleles. Expected bands for C57BL/6J, BALB/c and *Pcdh21*/Cre-Z/EG are 147, 213 and 100 bp, respectively. Primer sequences are: *D12Mit136* sense: 5'-TTTAATTTTGA GTGGGTTTGGC-3', antisense: 5'-TGGCTACATGTACTGATCTCCA-3'. PCR conditions were 94 °C for 2 min, 43 cycles of 94 °C for 1 min, 53 °C for 15 s, 72 °C for 45 s.

Albino allele PCR assay. Tetraploid blastocysts carry the C57BL/6J-*Tyr^{c-2j}* albino mutation, whereas the iMZ and iNZ iPS cells do not. We designed a PCR assay in which the 3' end of the sense primer is specific for the *Tyr^{c-2j}* mutation. C57BL/6J-*Tyr^{c-2j}* DNA yields the expected 115-bp product, whereas no product is observed with iMZ-9 DNA. DNA was collected from *Pcdh21*/Cre-Z/EG control and iPS mouse tissue by proteinase K digestion followed by phenol-chloroform extraction and ethanol precipitation. Primers used were: sense 5'-TCAAAGGGGTGGATGACCT-3', antisense 5'-CCCCCAAATCCA AACTTACA-3'. PCR conditions were 94 °C for 2 min, 40 cycles of 94 °C for 1 min, 65 °C for 15 s, 72 °C for 20 s.

qRT-PCR. RNA was collected from iPS cell lines maintained in the absence of dox (–dox) or treated with 10 μ g ml⁻¹ dox for 24 h to re-induce proviral encoded transgenes. As a control, RNA was collected from transiently transfected HEK293T cells expressing the individual lentiviruses. Total RNA was isolated using TRIzol reagent (Invitrogen), treated with DNase I and purified (RNeasy Plus kit, Qiagen) before synthesis of first-strand complementary DNA using the

SuperScript III First-Strand Synthesis System (Invitrogen). Quantitative PCR was performed on three technical replicates using the reverse transcription-real time SYBR green PCR Mastermix (SA Biosciences) and primers that specifically amplify the proviral encoded transgenes. Lentiviral-specific primers consist of a gene-specific sense primer and a common antisense primer located downstream of each transgene within the proviral backbone. Sense primers: *Oct4*, 5'-TCTGTTCCCGTCACTGCTCT-3'; *Sox2*, 5'-CGCCCAGTAGACTGCACAT-3'; *c-Myc*, 5'-TGTCCATTCAAGCAGACGAG-3'; *Klf4*, 5'-CACTACCGCAAACACACAGG-3'. Common antisense primer, 5'-GGCATTAAAGCAGCGTATCC-3'. PCR conditions were 94 °C for 4 min, 40 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s. Data was generated on a MJ Research Chromo4 PTC-200 thermal cycler and extracted with Opticon Monitor software.

Transgene expression level for iPS cells was normalized to *Gapdh* expression (*Gapdh* forward, 5'-TCAACGGGAAGCCCATCA-3'; *Gapdh* reverse, 5'-CTCGTGGTTCACACCCATCA-3') and plotted relative to transgenes expressed in transfected HEKs. The *Gapdh* primer pair did not amplify human *GAPDH* efficiently; therefore we normalized HEK293T transgene expression values to the average *Gapdh* expression value for iPS cells. It is important to note that although this analysis produces accurate relative expression levels for the same gene across various iPS cell lines, it provides only a rough estimate of the relative expression levels between transfected HEK293T cells and the iPS cells and this should not be considered quantitative.

When lentiviral expression was re-induced with dox, iPS cell lines tended to have one order of magnitude higher expression levels, indicating that the *rtTAM2.2* proviral insertion was not completely silenced in the iPS cell lines

and suggesting, by inference, that using a dox-inducible system can result in less residual transgene expression than non-inducible lentiviral strategies.

Generation of chimaeras. Chimaeras were produced by injection of iPS cells (passage 5–8) into diploid blastocysts, generated by mating superovulated C57BL/6J females to C57BL/6J × DBA2 F₁ stud males, according to the standard protocol²⁹.

Generation of iPS fetuses and mice. For tetraploid complementation, superovulated albino (BALB/cByJ × C57BL/6J-*Tyr*^{*c-2J*}/J) F₁ females were mated with males of the same hybrid strain background. One-cell embryos were collected and cultured overnight in KSOM-AA medium (Millipore). The next day tetraploid embryos were generated by blastomere electrofusion of two-cell embryos according to standard procedures and cultured under the same conditions^{15,29}. Forty-eight hours later, tetraploid blastocysts were injected with 10–12 iPS cells each and transferred to the uterine horns of pseudopregnant recipients. iPS-cell-derived fetuses were dissected from the uterine horns of recipient mice at different stages of development or live newborn pups were recovered by Caesarean section at E18.5 and fostered to CD-1 female mice.

30. Goebbels, S. *et al.* Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *Genesis* **44**, 611–621 (2006).
31. Wu, S. X. *et al.* Pyramidal neurons of upper cortical layers generated by NEX-positive progenitor cells in the subventricular zone. *Proc. Natl Acad. Sci. USA* **102**, 17172–17177 (2005).
32. Xu, J., Mashimo, T. & Sudhof, T. C. Synaptotagmin-1, -2, and -9: Ca²⁺ sensors for fast release that specify distinct presynaptic properties in subsets of neurons. *Neuron* **54**, 567–581 (2007).

Reptilian heart development and the molecular basis of cardiac chamber evolution

Kazuko Koshiba-Takeuchi^{1,2,3,4*}, Alessandro D. Mori^{1,2,3,5,6*}, Bogac L. Kaynak^{1,2,3*}, Judith Cebra-Thomas⁷, Tatyana Sukonnik^{1,2,3}, Romain O. Georges⁸, Stephany Latham⁹, Laural Beck⁹, R. Mark Henkelman^{10,11}, Brian L. Black^{3,12}, Eric N. Olson¹³, Juli Wade⁹, Jun K. Takeuchi⁴, Mona Nemer^{8,14}, Scott F. Gilbert¹⁵ & Benoit G. Bruneau^{1,2,3,5,6}

The emergence of terrestrial life witnessed the need for more sophisticated circulatory systems. This has evolved in birds, mammals and crocodilians into complete septation of the heart into left and right sides, allowing separate pulmonary and systemic circulatory systems, a key requirement for the evolution of endothermy^{1–3}. However, the evolution of the amniote heart is poorly understood. Reptilian hearts have been the subject of debate in the context of the evolution of cardiac septation: do they possess a single ventricular chamber or two incompletely septated ventricles^{4–7}? Here we examine heart development in the red-eared slider turtle, *Trachemys scripta elegans* (a chelonian), and the green anole, *Anolis carolinensis* (a squamate), focusing on gene expression in the developing ventricles. Both reptiles initially form a ventricular chamber that homogeneously expresses the T-box transcription factor gene *Tbx5*. In contrast, in birds and mammals, *Tbx5* is restricted to left ventricle precursors^{8,9}. In later stages, *Tbx5* expression in the turtle (but not anole) heart is gradually restricted to a distinct left ventricle, forming a left–right gradient. This suggests that *Tbx5* expression was refined during evolution to pattern the ventricles. In support of this hypothesis, we show that loss of *Tbx5* in the mouse ventricle results in a single chamber lacking distinct identity, indicating a requirement for *Tbx5* in septation. Importantly, misexpression of *Tbx5* throughout the developing myocardium to mimic the reptilian expression pattern also results in a single mispatterned ventricular chamber lacking septation. Thus ventricular septation is established by a steep and correctly positioned *Tbx5* gradient. Our findings provide a molecular mechanism for the evolution of the amniote ventricle, and support the concept that altered expression of developmental regulators is a key mechanism of vertebrate evolution.

Amphibians have a three-chambered heart, whereas mammalian, crocodilian and avian hearts have four chambers, two each for pulmonary and systemic circulations. The acquisition of a fully septated ventricle has evolved independently in birds, mammals and crocodilians¹⁰, and is an important example of convergent evolution. Non-crocodilian reptiles (squamates, chelonians and rhynchocephalians) hold a unique place in the evolution of the heart, as their ventricular chambers are apparent intermediates between these forms^{4–7}. In reptiles, shunting can produce functional separations between left and right circulatory

systems, but only complete septation allows a dual pressure system required for endothermy. Therefore, the evolutionary status of the reptilian ventricles is controversial⁷. Is it a primitive arrangement presaging the septated heart of crocodilians, birds and mammals? Or is it an adaptation to particular circulatory requirements? Development of reptilian hearts has not been addressed in over 100 years¹¹, and thus the developmental basis of reptilian heart formation is not known. Furthermore, clear insight into the evolution of cardiac septation has not emerged from molecular studies of heart development³.

Transcription factors of the T-box family are important regulators of heart formation¹². One T-box gene, *Tbx5*, has an expression pattern that suggests a role in the evolution of cardiac septation (see Supplementary Note 1). In amphibians, *Tbx5* is expressed throughout the developing heart¹³. In birds and mammals, there is a steep gradient of *Tbx5* expression from high levels in the prospective left ventricle to low levels in the prospective right ventricle^{8,9}. Reduced dosage of *Tbx5* in humans and mice leads to defects in interventricular septum (IVS) formation and patterning^{14–17}, suggesting that a steep gradient of *Tbx5* is critical for IVS formation. The evolutionary role of *Tbx5* in septation is unknown.

We examined cardiac embryology of the red-eared slider turtle, *T. scripta elegans* (a chelonian), and the green anole, *A. carolinensis* (a squamate), focusing on the ventricles. Although the phylogenetic relationship of turtles to other reptiles is controversial based on anatomical considerations^{18,19}, molecular phylogenies consistently group turtles with the archosaurs (birds and crocodiles)^{20,21}. Anoles are considered to be more basal than archosaurs^{19–21}. The post-hatching anole heart has a thick muscular ridge (Fig. 1a–d and Supplementary Fig. 1) that separates a proximal outflow tract, or cavum pulmonale^{6,11}, from the main ventricular chamber. Turtles have a smaller muscular ridge and are thought to have a primitive IVS-like structure^{4,6,11}, as we determined by three-dimensional reconstructions revealing a dense coalescence of trabeculae spanning the full depth of the heart (Fig. 1e–h and Supplementary Fig. 1). Initially, developing turtle and anole hearts showed no clear evidence of ventricular septation (Fig. 1i and Supplementary Figs 2–4). In contrast, the chick has a well-developed IVS at comparable early stages (Fig. 1i and Supplementary Fig. 3). In the turtle, a structure resembling an IVS appears only at stage 21 (Fig. 1i). Alligator

¹Gladstone Institute of Cardiovascular Disease, San Francisco, California 94158, USA. ²Department of Pediatrics, ³Cardiovascular Research Institute, University of California, San Francisco, California 94158, USA. ⁴Division of Cardiovascular Research, Global-Edge Institute, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan. ⁵Program in Stem Cell and Developmental Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁶Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁷Biology Department, Millersville University, Millersville, Pennsylvania 17551, USA. ⁸Institut de Recherches Cliniques de Montréal, Programme de Biologie Moléculaire, Université de Montréal, Montréal, Québec H3C 3J7, Canada. ⁹Department of Psychology and Program in Neuroscience, Michigan State University, East Lansing, Michigan 48824, USA. ¹⁰The Mouse Imaging Centre, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹¹Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ¹²Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. ¹³Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ¹⁴Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario K1H 8M5 Canada. ¹⁵Department of Biology, Swarthmore College, Swarthmore, Pennsylvania 19081, USA.

*These authors contributed equally to this work.

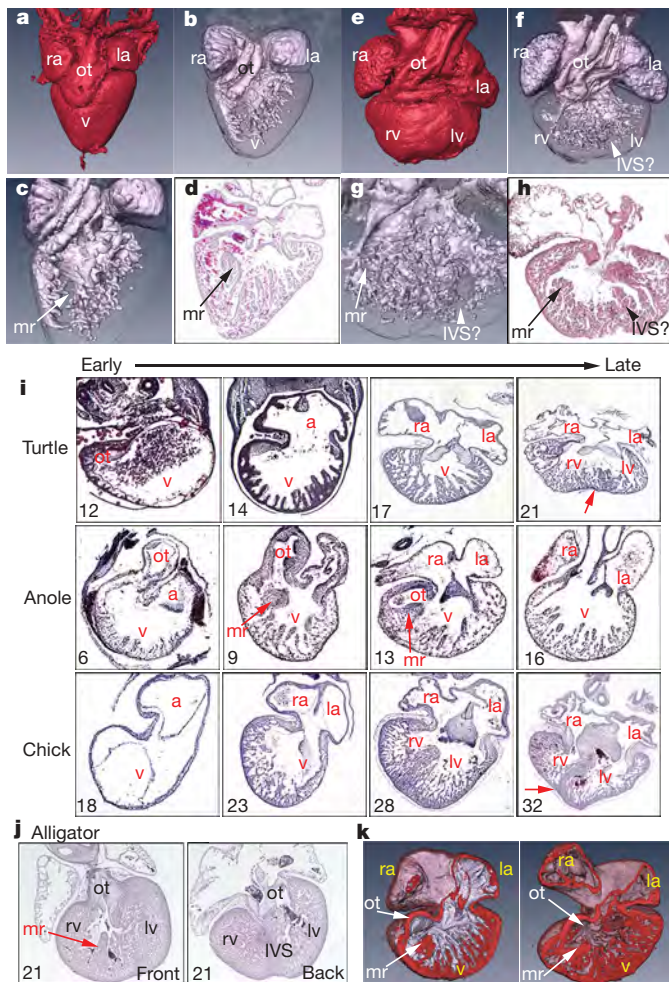


Figure 1 | Reptilian heart development. a–h, OPT of post-hatching anole (a–d) and turtle (e–h) hearts. a, e, External view; b, c, f, g, chamber fill; d, h, histology. i, Histological analysis of heart development in turtle, anole and chick embryos. Four representative stages shown are equivalent between species (arrow, interventricular groove). j, Histology of stage 21 embryonic alligator heart. k, OPT of stage 17 (left) and stage 21 (right) turtle hearts. In all reptile embryos (i–k), note close apposition of mr and ot. a, atrium; IVS?, IVS-like structure; la, left atrium; lv, left ventricle; mr, muscular ridge; ot, outflow tract; ra, right atrium; rv, right ventricle.

embryos (Fig. 1j and Supplementary Fig. 1) have a muscular ridge and a distinct IVS. The muscular ridge has been interpreted as analogous to the IVS, leading to the impression that reptiles have multiple septa^{4–6}. We speculate that the development of the muscular ridge in reptiles reflects persistent growth of the proximal outflow tract¹¹, as seen transiently in the chick heart (Fig. 1i–k and Supplementary Fig. 3).

To observe molecular patterning of reptile ventricles, we examined expression of *Tbx5*. In mammals and birds, *Tbx5* messenger RNA (mRNA) and protein are highly enriched in the prospective left ventricle (Fig. 2b, c and Supplementary Fig. 5)^{8,9}. At looping heart tube stages, *Tbx5* was broadly expressed throughout the embryonic turtle and anole hearts (Fig. 2a, d), similar to *Xenopus Tbx5* (ref. 13), but unlike its early restricted expression in chick and mouse (Fig. 2b, c). In the anole, *Tbx5* expression extended to the boundary of the ventricle and outflow tract, where the muscular ridge forms. At later stages, *Tbx5* expression in the turtle (stage 15) and anole (stage 13) remained homogeneous throughout the ventricle (Fig. 2e, h and data not shown). In comparable stages in chick, it was sharply restricted to left ventricle primordium. At stages 17–18 in the turtle, *Tbx5* mRNA levels decreased in right ventricle primordium, remaining enriched in left ventricle primordium, creating a steep left–right gradient,

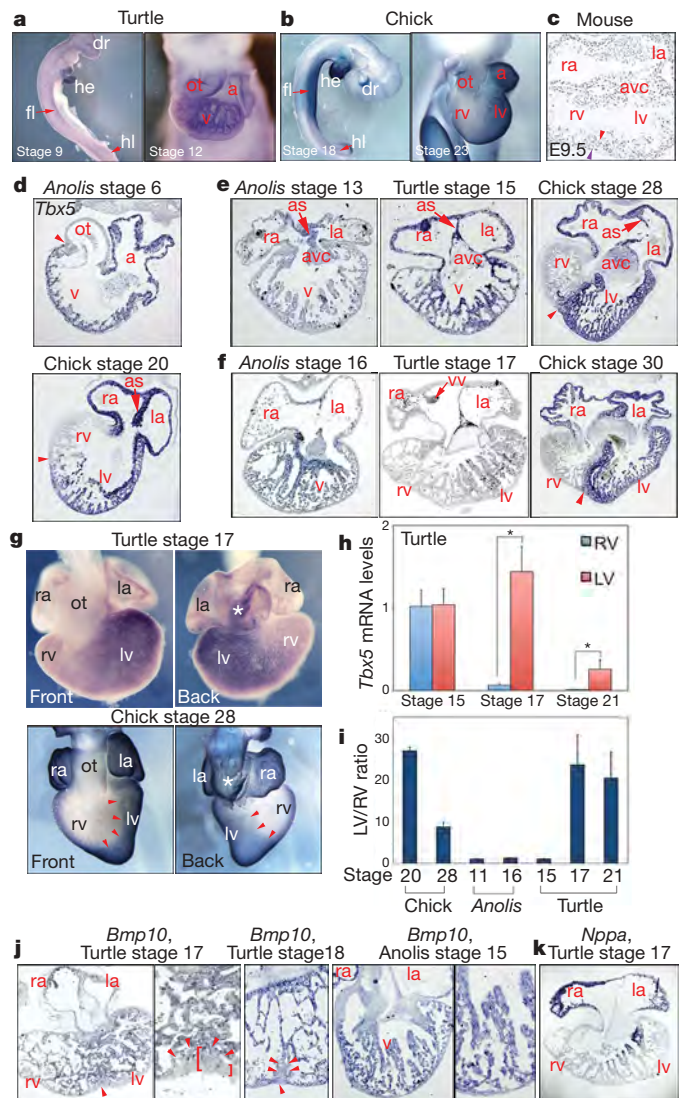


Figure 2 | Gene expression in amniote embryos. a–c, *Tbx5* expression in turtle, chick and mouse. a, b, mRNA expression. Left panels, whole-embryo views. he, heart; fl, forelimbs; dr, dorsal retina; hl, hindlimbs. Right panels, close-up ventral views of embryonic hearts. c, *Tbx5* immunohistochemistry; red arrowheads, right ventricle/left ventricle junction; purple arrowhead, epicardium. la, left atrium; lv, left ventricle; ot, outflow tract; ra, right atrium; rv, right ventricle. d–g, Expression of *Tbx5*. as, atrial septum; avc, atrioventricular cushion. Arrowheads mark the boundary between left ventricle and right ventricle, or ventricle (v) and ot for anole in d, h. Quantification of *Tbx5* mRNA levels in turtle left ventricle and right ventricle; data are mean \pm s.d. normalized to stage 15 right ventricle. * $P < 0.005$ by *t*-test. i, Ratio of *Tbx5* mRNA levels between the left ventricle and right ventricle. j, *Bmp10* expression in turtle and anole hearts. Arrowheads, interventricular groove and septum. Brackets, thickness of *Bmp10*-negative area. k, *Nppa* expression in the turtle is in a left–right gradient similar to *Tbx5*.

although not as sharply defined as in chick (Fig. 2f–i and Supplementary Figs 6 and 7). This gradient was maintained at stage 21 (Fig. 2h, i). *Tbx5* expression in *Anolis* was not restricted in the ventricle (Fig. 2f, i and Supplementary Fig. 7). We examined expression of *Tbx5* target genes expressed in trabeculae but excluded from mammalian IVS myocardium^{14,16,17,22}. *Bmp10* was expressed throughout the early turtle and anole trabeculae, but was excluded in turtles at stages 17–18 from an expansion of the compact myocardium corresponding to presumptive IVS precursors, correlating with the boundary of *Tbx5* expression (Fig. 2j and Supplementary Fig. 6). This suggests a conserved molecular transition in the trabeculae that form the IVS. Turtle *Nppa* (not found in anoles²³) formed a gradient

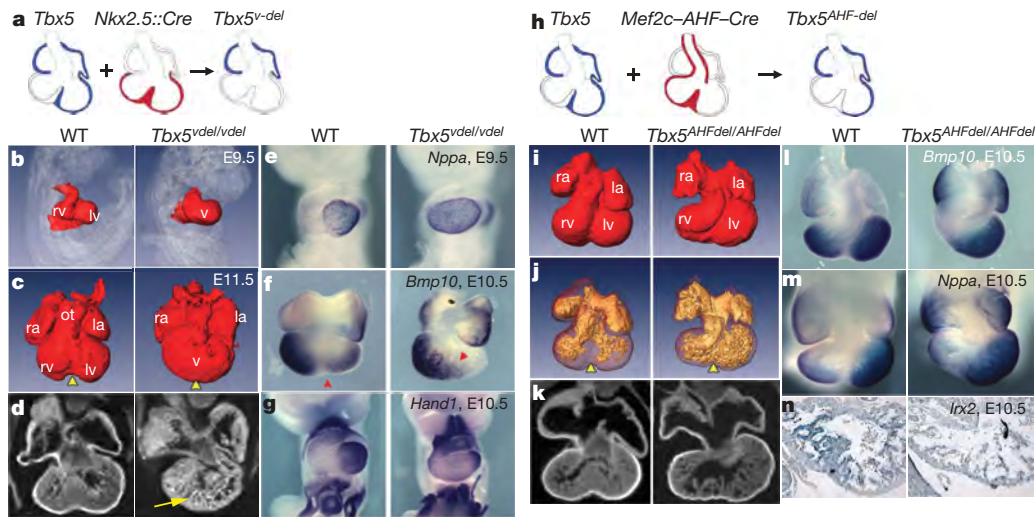


Figure 3 | Ventricular-restricted deletion of mouse *Tbx5*. **a**, Strategy for ventricular deletion of *Tbx5*. **b–d**, OPT of wild-type (WT) and *Tbx5*^{vdel/vdel} embryos and hearts at embryonic day (E) 9.5 (**b**) and E11.5 (**c**, **d**). Arrowheads indicate position of the IVS. **e–g**, Gene expression for indicated transcripts.

h, Strategy for *Tbx5* deletion in anterior heart field derivatives. **i–k**, OPT of wild-type (WT) and *Tbx5*^{AHFdel/AHFdel} hearts at E10.5. **i**, External view; **j**, chamber fill; **k**, virtual sections. **l–n**, Gene expression for indicated transcripts.

similar to *Tbx5* (Fig. 2k). Thus turtle ventricles, but not those of *Anolis*, acquire distinctions between left and right components late in development.

A steep *Tbx5* gradient in chick and mouse may have evolved to pattern the ventricles. Reducing *Tbx5* levels supports this^{14,16,17}. To

address a potential role for *Tbx5* in septation, we deleted *Tbx5* from segments of developing mouse ventricles, using a conditionally deletable *Tbx5* allele (*Tbx5*^{LDN})¹⁶, and ventricular myocyte-specific *Nkx2.5::Cre* mice²⁴ (Fig. 3a). These mice (*Nkx2.5::Cre*^{tg/tg}; *Tbx5*^{LDN/LDN} mice, or *Tbx5*^{vdel} mice) lacked morphological distinctions between the left ventricle and right ventricle that were obvious in wild-type embryos by embryonic day 9.5 (Fig. 3b). Embryos with this univentricular phenotype persisted until embryonic day 11.5 (Fig. 3c, d). Expression of *Nppa* and *Bmp10*, normally excluded from the interventricular groove, was expanded throughout the single ventricle of *Tbx5*^{vdel} embryos (Fig. 3e, f). *Hand1* was expressed at lower levels, but in its normal domains, the left ventricle and right ventricle primordia (Fig. 3g). Thus loss of *Tbx5* from developing ventricles results in a single mispatterned ventricle.

To determine if a steep *Tbx5* gradient at the interventricular midpoint is critical for IVS formation, we deleted *Tbx5* with *Mef2cAHF::Cre* mice²⁵ (Fig. 3h). Because *Mef2cAHF::Cre* is active in the right ventricle and IVS precursors, but not in the left ventricle free wall, the *Tbx5* expression boundary is shifted leftwards. *Tbx5*^{LDN/LDN}; *Mef2cAHF::Cre* (*Tbx5*^{AHFdel}) mice lacked an IVS (Fig. 3i–k). Gene expression analysis showed that a distinction between left ventricle and right ventricle was maintained (Fig. 3l, m); however, a clear absence of IVS-enriched markers (*Irx2*, *Dkk3*) at the ventricular midpoint, although maintained in the adjacent trabeculae, emphasizes the absence of ventricular septation (Fig. 3n and Supplementary Fig. 8). Thus a boundary of cells expressing high *Tbx5* levels is necessary within a segment of myocardium where IVS outgrowth will occur. This implies a prepattern within which *Tbx5* must function; the nature of this prepattern is unknown (see Supplementary Note 2). *Tbx5* expression and additional patterning cues may have co-evolved, or the prepattern may exist in all amniotes. Regardless, IVS formation requires a sharp *Tbx5* boundary, which indicates that *Tbx5* patterning was a major factor in the evolution of septation.

Our loss-of-function experiments demonstrate a requirement for *Tbx5* in IVS formation distinct from a more global role in differentiation. These results do not address the evolutionary role of *Tbx5* patterning; in particular, whether the broad expression of *Tbx5* observed in anole and turtle would preclude IVS formation. Previous misexpression attempts yielded variable results ranging from no effect to severely malformed hearts (ref. 9 and J.K.T., unpublished observations). We misexpressed *Tbx5* in the ventricles by crossing a mouse line bearing a stable Cre-activatable transgene expressing moderate

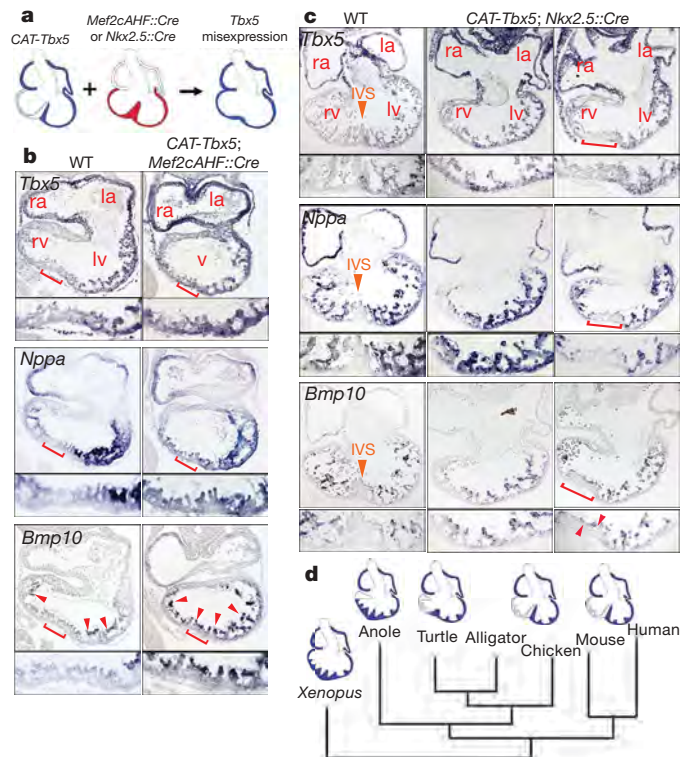


Figure 4 | Misexpression of *Tbx5* results in loss of IVS patterning.

a, Strategy for ventricular misexpression of *Tbx5*. **b**, Morphology and gene expression in *CAT-Tbx5*; *Mef2cAHF::Cre* embryos for indicated transcripts. Brackets, IVS region, magnified in lower panels. Arrowheads, trabecular *Bmp10* expression. la, left atrium; lv, left ventricle; ra, right atrium; rv, right ventricle; v, ventricle. **c**, Morphology and gene expression in *CAT-Tbx5*; *Nkx2.5::Cre* embryos at embryonic day 11.5. Orange arrows, interventricular septal region (IVS). Brackets show a rudimentary septum in a mutant embryo. **d**, Representation of embryonic heart structures and patterns of *Tbx5* expression (blue) in vertebrate evolution.

Tbx5 levels upon induction (*CAT-Tbx5*)²⁶ with *Mef2cAHF::Cre* or *Nkx2.5::Cre* mice (Fig. 4). *CAT-Tbx5;Mef2cAHF::Cre* embryos survived until embryonic day 11 and had a single ventricle at embryonic day 10.25. Molecular analysis revealed expanded expression of *Tbx5*, *Nppa* and *Bmp10* across the interventricular groove of *CAT-Tbx5;Mef2cAHF::Cre* embryos (Fig. 4b). *CAT-Tbx5;Nkx2.5::Cre* embryos survived longer (until embryonic day 12), presumably because this manipulation avoided secondary effects of Tbx5 overexpression in cardiac progenitors. *CAT-Tbx5;Nkx2.5::Cre* embryos at embryonic day 11.5 also had defective ventricular septation and mispatterned gene expression (Fig. 4c and Supplementary Fig. 9). Interestingly, owing to the mosaic expression of Tbx5 by *Nkx2.5::Cre*, some embryos had no septum at all, whereas others with a more graded expression of Tbx5 had a rudiment of a septum in which not all genes were mispatterned (Fig. 4c and Supplementary Fig. 9). Thus misexpression of Tbx5 in a pattern reminiscent of the reptilian heart leads to loss of IVS patterning and morphogenesis, further supporting a role for Tbx5 patterning in the evolution of septation.

Our results provide evidence that the reptilian heart, although evolved to function physiologically under conditions particular to reptilian life⁷, is an evolutionary intermediate between amphibian and avian/crocodilian hearts in its ventricular development. The dynamic expression of *Tbx5* and its leftward restriction suggest a temporal refinement model in which early restriction of *Tbx5* expression to left ventricle precursors, as seen in chick and mouse, provides a robust patterning cue for ventricular septation. In this model (Fig. 4d), a quantitative gradient of Tbx5 is essential for proper formation and patterning of the IVS. Our mouse genetic analyses, including decreased dosage^{14,16}, are consistent with an important role for a steep gradient of Tbx5 in chamber patterning and IVS formation. In the reptilian heart, the delayed and less pronounced establishment of this patterning may contribute to varying degrees of septation. Therefore patterning of *Tbx5*, in the archosaurian and synapsid lineages, is likely to be an important mechanism in the convergent evolution of septation. Our findings generally support the concept that altered expression of developmental regulators is an important aspect of morphological evolution²⁷.

METHODS SUMMARY

Embryos were isolated from *T. scripta elegans* eggs (Kliebert Turtle and Alligator Farm). Green anole (*A. carolinensis*) embryos were collected in captivity. Mouse strains were as previously described^{16,24–26}. Wholemount and section *in situ* hybridizations were performed using standard protocols. Immunohistochemistry and optical projection tomography (OPT) were performed as previously described^{26,28}. For all mouse experiments, at least three embryos were examined for each genotype at each stage, all with comparable results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 December 2008; accepted 27 July 2009.

- Farmer, C. G. Evolution of the vertebrate cardio-pulmonary system. *Annu. Rev. Physiol.* **61**, 573–592 (1999).
- Hillenius, W. J. & Ruben, J. A. The evolution of endothermy in terrestrial vertebrates: Who? When? Why? *Physiol. Biochem. Zool.* **77**, 1019–1042 (2004).
- Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
- Holmes, E. B. A reconsideration of the phylogeny of the tetrapod heart. *J. Morphol.* **147**, 209–228 (1975).
- Webb, G. J. W. Comparative cardiac anatomy of the Reptilia. III. The heart of crocodilians and an hypothesis on the completion of the interventricular septum of crocodilians and birds. *J. Morphol.* **161**, 221–240 (1979).
- Farrell, A. P., Gamperl, A. K. & Francis, E. T. B. In *Biology of the Reptilia* Vol. 19 (eds Gans, C. & Gaunt, A. S.) 375–424 (Society for the Study of Amphibians and Reptiles, 1998).
- Hicks, J. W. The physiological and evolutionary significance of cardiovascular shunting patterns in reptiles. *News Physiol. Sci.* **17**, 241–245 (2002).
- Bruneau, B. G. *et al.* Chamber-specific cardiac expression of Tbx5 and heart defects in Holt–Oram syndrome. *Dev. Biol.* **211**, 100–108 (1999).
- Takeuchi, J. K. *et al.* Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. *Development* **130**, 5953–5964 (2003).

- Seymour, R. S. *et al.* Evidence for endothermic ancestors of crocodiles at the stem of archosaur evolution. *Physiol. Biochem. Zool.* **77**, 1051–1067 (2004).
- Greil, A. Beiträge zur vergleichenden Anatomie und Entwicklungsgeschichte des Herzens und des Truncus arteriosus der Wirbeltiere. *Morphol. Jahrb.* **31**, 123–310 (1903).
- Stennard, F. A. & Harvey, R. P. T-box transcription factors and their roles in regulatory hierarchies in the developing heart. *Development* **132**, 4897–4910 (2005).
- Horb, M. E. & Thomsen, G. H. Tbx5 is essential for heart development. *Development* **126**, 1739–1751 (1999).
- Bruneau, B. G. *et al.* A murine model of Holt–Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).
- Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
- Mori, A. D. *et al.* Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev. Biol.* **297**, 566–586 (2006).
- Koshiba-Takeuchi, K. *et al.* Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nature Genet.* **38**, 175–183 (2006).
- Rieppel, O. Turtle origins. *Science* **283**, 945–946 (1999).
- Lyson, T. & Gilbert, S. F. Turtles all the way down: loggerheads at the root of the chelonian tree. *Evol. Dev.* **11**, 133–135 (2009).
- Hedges, S. B. & Poling, L. L. A molecular phylogeny of reptiles. *Science* **283**, 998–1001 (1999).
- Shedlock, A. M. *et al.* Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl Acad. Sci. USA* **104**, 2767–2772 (2007).
- Chen, H. *et al.* BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* **131**, 2219–2231 (2004).
- Trajanovska, S. & Donald, J. A. Molecular cloning of natriuretic peptides from the heart of reptiles: loss of ANP in diapsid reptiles and birds. *Gen. Comp. Endocrinol.* **156**, 339–346 (2008).
- McFadden, D. G. *et al.* The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189–201 (2004).
- Verzi, M. P. *et al.* The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev. Biol.* **287**, 437–449 (2005).
- Georges, R. *et al.* Distinct expression and function of alternatively spliced Tbx5 isoforms in cell growth and differentiation. *Mol. Cell. Biol.* **28**, 4052–4067 (2008).
- Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
- Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107–112 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. N. Wylie and L. Davidson for technical assistance, M. Harris and J. Fallon for alligator embryos, T. Sanger and J. Gibson-Brown for unpublished data on *Anolis* staging, T. Ogura for chick *Tbx5* and *Tbx20* probes, and G. Howard and S. Ordway for editorial assistance. This work was funded in part by the March of Dimes Birth Defects Foundation (B.G.B.), the J. David Gladstone Institutes (B.G.B.), William H. Younger Jr (B.G.B.); a National Institutes of Health program project grant (P01HL089707 to B.G.B., B.L.B.); scholarships from the Natural Sciences and Engineering Research Council of Canada, the Heart and Stroke Richard Lewar Centre for Excellence, University of Toronto, and Ontario Graduate Scholarship (A.D.M.); the Fumi Yamamura Memorial Foundation for Female Natural Scientists and Grants-in-Aid for Scientific Research (C) (K.K.-T.), MEXT's program for young independent researchers (K.K.-T., J.K.T.), Sumitomo Foundation and Nakajima Foundation (J.K.T.), a Canada Research Chair in Imaging (R.M.H.), the Heart and Stroke foundation of Canada and the Canadian Institutes for Health Research (M.N.), and grants from the National Science Foundation (RUI-0748508 to S.F.G., J.C.-T. and IOS-0742833 to J.W.). Funding for the J. David Gladstone Institutes from a National Institutes of Health/ National Center for Research Resources grant (CO6 RR018928) is also acknowledged.

Author Contributions K.K.-T. performed reptile histology and gene expression studies; A.D.M., B.L.K., T.S. and B.G.B. performed mouse experiments; J.C.-T. and S.F.G. obtained turtle specimens and isolated *T. scripta* *Tbx5* complementary DNA (cDNA); S.L. and L.B. isolated *Anolis* specimens under direction of J.W.; B.L.K. acquired and reconstructed OPT images; R.O.G. performed Tbx5 immunohistochemistry under direction of M.N.; R.M.H. directed initial mouse embryo OPT; J.K.T. obtained chick and mouse specimens; R.O.G., M.N., B.L.B. and E.N.O. provided genetically modified mice before publication; B.G.B. conceived and directed the project, and wrote the paper. All authors contributed to the written manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.G.B. (bbruneau@gladstone.ucsf.edu).

METHODS

Reptiles. Embryos from *T. scripta elegans* eggs (Kliebert Turtle and Alligator Farm) were dissected free of extra-embryonic membranes and staged according to Greenbaum & Carr²⁹. Green anole (*A. carolinensis*) embryos were collected in captivity during the spring and summer breeding season. The day the eggs were discovered in the cage was considered embryonic day 0. Eggs were incubated at 27 °C until embryo collection at embryonic days 4, 8, 11, 12, 14, 18 and 19 and on the day of hatching. Anole embryos were staged as previously described³⁰. All embryos were fixed in 4% paraformaldehyde, followed by stepwise dehydration in methanol.

Mouse genetics. Mouse strains were as previously described^{16,24–26}.

Reptile cDNAs. Turtle total RNA was prepared from pooled day 14 *T. scripta* embryos (stages 12–15). PCR was performed from cDNA with 5'-GTTTC CCAGTTACAAAGTGAAGG (forward) and 5'-GTCTCACTGTGCTCCTGGG (reverse) primers designed against the chick *Tbx5* sequence. PCR products

matching the estimated size (540 base pairs) were extracted from a 1% agarose gel, ligated into pCRII (Invitrogen) and used to transform DH5 α -competent cells. Positive plasmids were identified by hybridization with digoxigenin-labelled chick *Tbx5* and by amplification using nested primers also designed against chick *Tbx5* and *Tbx4* (5'-TAYGTGCACCCGGAYTCYCCWGC and 5'-TGGTASGARGTCACAGMGATRAA). *Anolis* cDNA probes were obtained by PCR from genomic DNA, using sequences obtained at the National Center for Biotechnology Information Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>).

29. Greenbaum, E. & Carr, J. L. Staging criteria for embryos of the spiny softshell turtle, *Apalone spinifera* (Testudines: Trionychidae). *J. Morphol.* **254**, 272–291 (2002).
30. Sanger, T. J., Losos, J. B. & Gibson-Brown, J. J. A developmental staging series for the lizard genus *Anolis*: a new system for the integration of evolution, development, and ecology. *J. Morphol.* **269**, 129–137 (2008).

Coordination of Rho GTPase activities during cell protrusion

Matthias Machacek^{1*†}, Louis Hodgson^{2*†}, Christopher Welch^{2*}, Hunter Elliott¹, Olivier Pertz^{1†}, Perihan Nalbant³, Amy Abell², Gary L. Johnson², Klaus M. Hahn² & Gaudenz Danuser¹

The GTPases Rac1, RhoA and Cdc42 act together to control cytoskeleton dynamics^{1–3}. Recent biosensor studies have shown that all three GTPases are activated at the front of migrating cells^{4–7}, and biochemical evidence suggests that they may regulate one another: Cdc42 can activate Rac1 (ref. 8), and Rac1 and RhoA are mutually inhibitory^{9–12}. However, their spatiotemporal coordination, at the seconds and single-micrometre dimensions typical of individual protrusion events, remains unknown. Here we examine GTPase coordination in mouse embryonic fibroblasts both through simultaneous visualization of two GTPase biosensors and using a ‘computational multiplexing’ approach capable of defining the relationships between multiple protein activities visualized in separate experiments. We found that RhoA is activated at the cell edge synchronous with edge advancement, whereas Cdc42 and Rac1 are activated 2 μ m behind the edge with a delay of 40 s. This indicates that Rac1 and RhoA operate antagonistically through spatial separation and precise timing, and that RhoA has a role in the initial events of protrusion, whereas Rac1 and Cdc42 activate pathways implicated in reinforcement and stabilization of newly expanded protrusions.

Our computational multiplexing approach makes the assumption that the relationship between GTPase activation and the movements of the cell edge during constitutive protrusion and retraction cycles is preserved within a cell, and among cells. Thus, the initiation of protrusion and retraction can be used as a timing reference to determine indirectly the activation dynamics of multiple Rho GTPases: first, for each GTPase, the timing of activation relative to protrusion/retraction events is determined in separate experiments. Then, the activation timings of different GTPases are aligned, using protrusion/retraction events as a common reference. GTPase activities were imaged in mouse embryonic fibroblasts (MEFs) using biosensors for Rac1 (ref. 4), Cdc42 (ref. 5) or RhoA⁶ (Fig. 1a–c and Supplementary Figs 1 and 2). The Rac1 biosensor was improved compared with the previously published version⁴, using a fluorescent protein rather than a covalently attached dye. Images were captured at 10-s intervals, sufficient to sample the protrusion–retraction cycles below Nyquist frequency (Supplementary Fig. 3).

As reported before^{4–6}, all GTPases, including RhoA, were maximally activated proximal to the leading edge. Visual inspection of time-lapse sequences indicated substantial fluctuations in GTPase activity as the cell edges protruded and retracted (Supplementary Movies 1–6). To quantify the magnitude and location of the fluctuations we extracted time courses of GTPase activity at multiple distances from the cell edge and quantified the extent of signal modulation (Fig. 1d). For all three

GTPases the modulation was highest at the cell edge, decreased monotonically within \sim 2–4 μ m, and then reached a plateau of baseline activity (Fig. 1e–g). Rac1 displayed the least significant decay, mainly because of cell-to-cell variations in the extent of the region with high activation. RhoA, on the other hand, decayed tightly over 2 μ m. Thus, GTPase activities are regulated most prominently within a few micrometres from the leading edge, supporting the hypothesis that these fluctuations may be linked to cell edge movements.

To investigate how GTPase activation relates to edge movement spatially and temporally, we tracked the position of the cell edge¹³ and compared edge velocities with the biosensor signal in 40 to 80 sampling windows moving with the leading edge (Fig. 1h and Supplementary Movie 7). The window width was set to 1.8–3 μ m, the distance over which edge movements were correlated along the cell boundary, and 0.9 μ m in depth, the distance over which signalling molecules bound to the plasma membrane would be expected to diffuse between consecutive movie frames¹⁴. Edge velocities were sampled every 400 nm (see Methods), so that 5–7 velocity measurements would fall within a window (Fig. 1i). Thus, each window yielded time courses of both edge velocity and GTPase activation level (Fig. 1j), with independent measurements in each adjacent window. Changes in edge velocity appeared to parallel changes in GTPase activity (shown for the example of a Cdc42 data set), but with a time lag.

To study more systematically the potential relationships between edge dynamics (Fig. 2a, d, g) and GTPase activation we copied the data from the individual sampling windows along the cell boundary into activity maps¹³ (Fig. 2b, c, e, f, h, i). This data representation revealed tight correlation of morphological dynamics and GTPase activation. Protrusion and retraction occurred in a quasi-cyclic fashion with a periodicity of \sim 100 s (Supplementary Fig. 3). As reported previously¹³, protrusion cycles propagated transversally along the leading cell edge (Fig. 2b, e, h). Both Cdc42 and RhoA cycled with similar periodicity (Supplementary Fig. 3) and their activity maps visually resembled those of the morphological dynamics (Fig. 2f, i). Rac1 showed a broader time response (Supplementary Fig. 3), consistent with a spatiotemporally more diffuse activation (Fig. 2c). Quiescent regions of the cell edge (that is, regions Q) exhibited low GTPase activity compared with protruding and retracting regions. Transitions between active and quiescent portions of the cell edge were sharp (2–3 μ m) and were accompanied by a steep gradient in GTPase activity. Thus, these data suggest an immediate coupling of GTPase activation and cell morphological dynamics, in both time and space.

To quantify the coupling we computed, for every sampling window, Pearson’s correlation coefficient between edge velocity

¹Department of Cell Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA. ²Departments of Pharmacology, Medicinal Chemistry and Lineberger Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³Department of Molecular Cell Biology, Center for Medical Biotechnology, University of Duisburg-Essen, 45117 Essen, Germany. [†]Present addresses: Novartis Pharma AG, Lichtstrasse 35, CH-4056 Basel, Switzerland (M.M.); Department of Anatomy and Structural Biology and Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Ave, Bronx, New York 10461, USA (L.H.); Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland (O.P.).

*These authors contributed equally to this work.

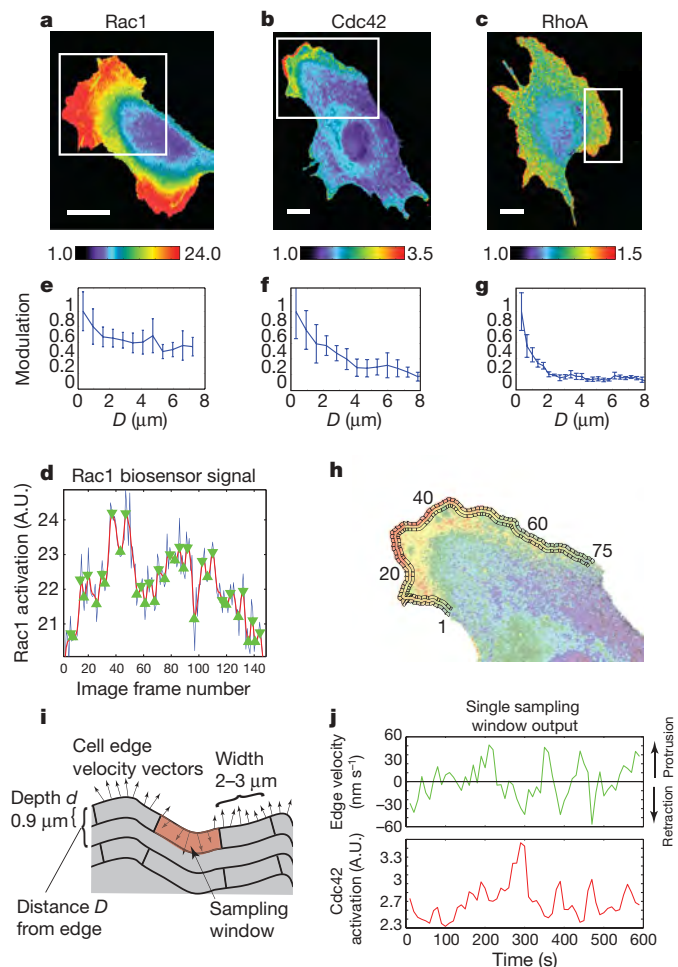


Figure 1 | Activation of Rho GTPases in migrating MEFs. **a–c**, Rac1 (**a**), Cdc42 (**b**) and RhoA (**c**) activation reported by biosensors. White box indicates region of interest selected for analysis. Scale bar, 20 μm . Colour scale indicates the dynamic range of the biosensor response (1, no significant response; maximum value, strongest response throughout the time-lapse sequence). **d**, Time course (blue line) of Rac1 activation within 0.6 μm from the cell edge, averaged over a 5- μm -long portion of the cell edge. Red line indicates filtered time course (Gaussian filter, $\sigma = 5$ frames). Green triangles indicate local maxima and minima of filtered time course. **e–g**, Modulation of GTPase activation (mean absolute difference between consecutive local extrema of the time course) as a function of the distance D from the leading edge. Values are normalized to the modulation at the cell edge. Error bars indicate s.e.m. of $n = 6$ (Rac1), $n = 4$ (Cdc42) and $n = 4$ (RhoA) cells. **h**, Sampling windows of 0.9- μm depth placed at the cell edge and at $D = 1.8 \mu\text{m}$ from the cell edge. **i**, Parameters to define the position and size of a sampling window. In each window, the time course of GTPase activation was recorded (average of ~ 10 pixels). For a sampling window at $D = 0$, a time course of edge velocity was recorded (mean of 6–8 measurements). **j**, Time courses of edge velocity (green) and GTPase activation (Cdc42, red) recorded in one sampling window.

and the GTPase activity, as a function of the time lag between the two variables. High correlation coefficients indicated a tight coupling of GTPase activation and cell protrusion/retraction events, whereas low correlation coefficients suggested a more remote relationship. Moreover, if a significant maximum correlation coefficient was obtained at a non-zero time lag, it indicated the relative timing of the two processes. For actively protruding regions, we observed significant correlation coefficients for all three GTPases. For Cdc42 and Rac1, high correlation values occurred predominantly at negative time lags (Fig. 2j, k), that is, the activation of these two GTPases was delayed relative to cell edge advancements. In contrast, for RhoA high correlation values were narrowly distributed around zero time lags (Fig. 2l). Thus, RhoA activation seemed to be synchronous

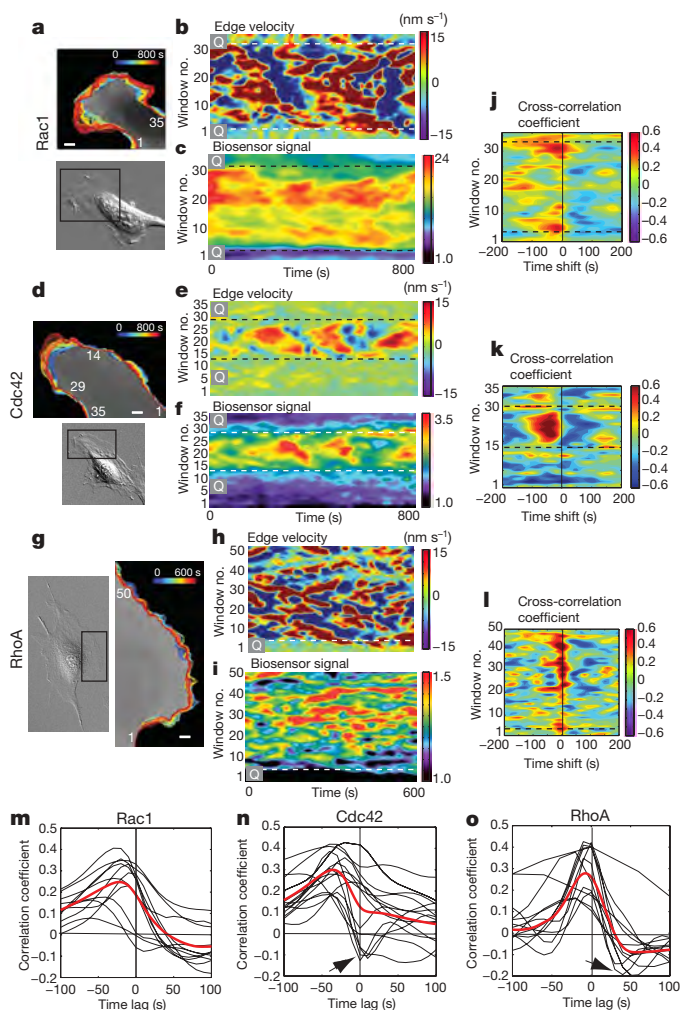


Figure 2 | Dynamics of cell edge morphology and GTPase activation. **a, d, g**, Overview of cell morphology (in DIC) and evolution of cell edge positions (colour-encoded from blue (early) to red (late) time points). Scale bar, 5 μm . **b, e, h**, Activity maps of edge movement: velocity values along the edge are copied time point by time point into the columns of the map. Colour coding: red, protrusion; blue, retraction. Quiescent regions of the cell edge are marked by 'Q'. **c, f, i**, Activity maps of Rho GTPase activation: colours designate the activation level of the Rho GTPase recorded in a sampling window at $D = 0 \mu\text{m}$. **j, k, l**, Cross-correlation coefficients (colour-coded) between edge velocity and Rac1 (**j**), Cdc42 (**k**) and RhoA (**l**) as a function of the sampling window and time shift. **m, n, o**, Temporal cross-correlation functions for individual cells (black). Red lines indicate average cross-correlation functions per GTPase resulting from a spline fit to the combined single-cell cross-correlation functions; $n = 11$ (Rac1), $n = 12$ (Cdc42) and $n = 12$ cells (RhoA). Some cells show negative correlation coefficients at positive time lags (arrowheads), probably associated with ruffling (see text).

with forward movements of the cell edge. These results show that GTPases are activated over a fixed time interval relative to the dynamics of the leading edge.

To estimate more precisely the activation of the three GTPases relative to protrusion/retraction events, we averaged the correlation coefficients over all sampling windows of a cell (Supplementary Fig. 4). A comparison between cells showed that despite substantial variation in shape and morphodynamic behaviour, the relationships between GTPase activation and edge movements were preserved (Fig. 2m–o). To estimate the average timing between protrusion/retraction and GTPase activation across cells the correlation coefficients of individual cells were pooled and the significant maximum of a spline function fitted through the data determined (Fig. 2m–o, red curves; Supplementary Methods). A few cells displayed correlation minima

at positive time lags, particularly with Cdc42 and RhoA (Fig. 2n, o, arrowheads). These were probably due to ruffling events in which GTPases reach a high activation level a few seconds before the lamellipodium is lifted off the substrate and the edge retracts (see Supplementary Movies 3–4). Edge retraction generates a negative velocity in the velocity map, and hence a negative cross-correlation with elevated GTPase activation. These ruffling events did not affect our finding that both Rac1 and Cdc42 activation lag edge protrusion, whereas RhoA is activated during cell protrusion.

We speculated that the GTPases might be organized not only temporally, but also spatially, within the band of modulated signalling activity (see Fig. 1e–g). Therefore, we repeated the correlation analysis between protrusion and retraction events and activation of Rac1, Cdc42 and RhoA in sampling windows at various distances from the cell edge (Fig. 3a–c, see also Supplementary Movie 7). Maxima of the correlation coefficients CM and the corresponding time lags Δt were determined for each distance and fitted with a smoothing spline (Fig. 3d–f). For RhoA the correlation coefficient was highest at the leading edge and monotonically decreased at larger distances from the edge (Fig. 3f, blue line). At $D > 2 \mu\text{m}$ the correlation coefficient was smaller than 0.2 (Fig. 3d–f, black dashed line), the 95% confidence level for correlation coefficients of individual cells (Supplementary Methods). Thus, beyond $2 \mu\text{m}$ RhoA activity is no longer related to protrusion events (Fig. 3f, arrow). Rac1 and Cdc42 activities were correlated with edge movements over a wider region ($4.5 \mu\text{m}$ and $3.2 \mu\text{m}$, respectively; Fig. 3d, e, blue curve). Interestingly, the highest correlation coefficients were found at a distance $D = 1.8 \pm 0.7 \mu\text{m}$ for Rac1 and at $D = 1.3 \pm 0.7 \mu\text{m}$ for Cdc42. The same distance D was obtained when identifying the distance with the shortest time lag between edge velocity fluctuations and GTPase activation (Fig. 3d, e, red curve), which indicates the location of initial GTPase activation. These results indicate that both Rac1 and Cdc42 are activated at $\sim 1.8 \mu\text{m}$ from the cell edge. The larger time lags and lower correlation coefficients at distances other than $1.8 \mu\text{m}$ suggest that the activation of Rac1 and Cdc42 then propagates in anterior and posterior directions, losing the close coordination with edge movements.

Next, we estimated the between-cell variation of the cross-correlation time lag (Fig. 4a) by bootstrapping 2,000 samples from the residual distribution of the spline fit to the individual correlation functions in Fig. 2m–o. The time lags were consistently negative for all three GTPases, that is, the peak in their activation was delayed relative to

the protrusion of the cell edge (time lag (s) to protrusion for Rac1, -41 (-61 -13); Cdc42, -46 (-51 -39); RhoA, -6 (-8 -4); mean \pm 95% confidence interval as determined by bootstrap analysis). Hence, Rac1, Cdc42 and RhoA each have a different timing of activation, which is conserved between cells.

We were concerned that the measured spatiotemporal shifts of Rac1, Cdc42 and RhoA activation might be attributed to the different biosensor designs we used for each of these proteins (Supplementary Fig. 1). Some designs might be more susceptible to competition from native ligands, affecting readouts of activation kinetics. Therefore, we titrated the concentrations of biosensor components (Supplementary Figs 8 and 9) and repeated the timing experiments using different biosensor designs for the same GTPase (Supplementary Fig. 10). These control experiments confirmed that neither the relative activation kinetics nor the localization were a function of biosensor design (see Supplementary Materials).

We also sought to validate the time shifts between GTPase activations predicted by computational multiplexing by directly observing two signalling activities in the same cell. To this end, we modified the Cdc42 biosensor with new fluorophores (Supplementary Fig. 6), enabling simultaneous imaging of Cdc42 and RhoA activation in one cell (Movie 9). The resulting four-channel image data provided unprecedented temporal and spatial resolution for the study of the coordination between two signalling molecules, without using the cell edge velocity as a reference. Here, fluctuation correlation analysis was valuable to harness the full information from multiplex biosensor imaging. The data were again analysed locally in sampling windows at the leading edge. Averaging the correlation coefficients between the Cdc42 and RhoA time courses within one cell and over $n = 7$ cells we obtained a time lag of -34 (-30 , -38) s, that is, Cdc42 was activated after RhoA (Fig. 4b). This was within the confidence band of the computationally predicted difference between Cdc42 and RhoA activation (see Fig. 4a), confirming that direct visualization and indirect inference of signalling relationships yield the same result.

Together, these data indicate the following dynamics of Rho family GTPase activation in one protrusion–retraction cycle (Fig. 4c): RhoA activation increases and decreases in synchrony with protrusion and retraction. RhoA activation that is correlated with leading edge dynamics is confined to a band $2 \mu\text{m}$ from the leading edge. Cdc42 and Rac1 reach their peak activation with a 40-s delay relative to protrusion, and their activation is initiated $1.8 \mu\text{m}$ from the leading edge. Rac1 and Cdc42 are temporally as well as spatially less coupled

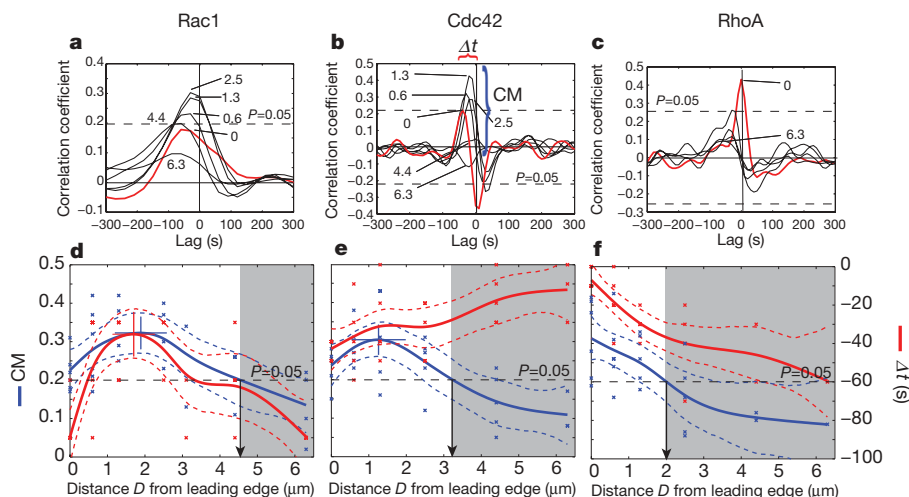


Figure 3 | Correlation of Rho GTPase activation and cell edge velocity as a function of time and space. a–c, Cross-correlation functions between edge velocity and GTPase activation at $D = 0$ (red line), 0.6, 1.3, 2.5, 4.4 and 6.3 μm . Dashed line indicates 95% confidence level for correlation values. d–f, Blue line indicates magnitude of cross-correlation maximum (CM has no units) as a function of D (blue bracket in b). Red line indicates time lag of

cross-correlation maxima as a function of D (red bracket in b). Dashed lines indicate 95% confidence interval estimated by bootstrap sampling of residual distributions to the spline fits (see Supplementary Methods). Grey area indicates region with correlation values below 95% confidence. Of note, time lags of cross-correlation maxima in this area are no longer meaningful. Data derived from $n = 6$ (Rac1), $n = 5$ (Cdc42) and $n = 5$ (RhoA) cells.

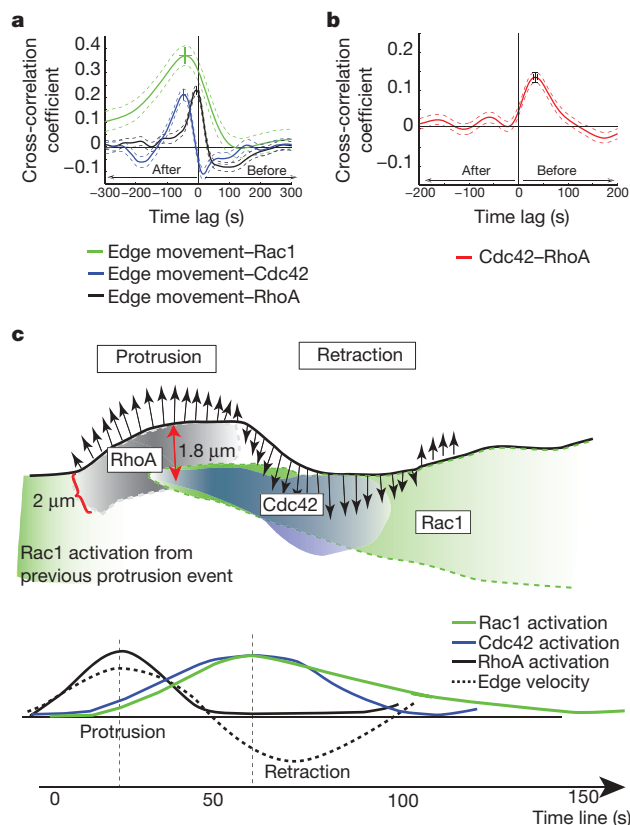


Figure 4 | Spatiotemporal coordination of Rac1, Cdc42 and RhoA activation. **a**, Timing of Rho GTPase activation relative to edge velocity, as determined by temporal cross-correlation functions. The variable 'edge movement' is being used as the reference signal. Thus, correlation maxima in the sector labelled 'After' indicate that the GTPase reaches the activation maximum after the protrusion event (time point of fastest edge advancement). Dashed lines indicate 95% confidence intervals. Data originate from $n = 11$ (Rac1), $n = 12$ (Cdc42) and $n = 12$ cells (RhoA). Confidence intervals were computed by bootstrap sampling of 2,000 residuals to the spline fits. **b**, Timing of RhoA activity relative to Cdc42 activity, both monitored in the same cell, as determined by the temporal cross-correlation function. The variable Cdc42 is being used as the reference signal. Thus, the correlation maximum in the sector labelled 'Before' indicates that RhoA reaches the activation maximum before Cdc42 ($n = 7$ cells). The y axes for panels **a** and **b** are without units. **c**, Model of GTPase activation during protrusion and retraction. Green indicates Rac1 activation; black, RhoA activation; blue, Cdc42 activation.

to protrusion. Cdc42, and even more so Rac1, maintain significant levels of activity during the retraction phase, such that lower levels of Rac1 activation overlap the onset of the next protrusion cycle. Rac1 and Cdc42 signals propagate from their location of initial activation in anterograde and retrograde directions via as-yet-unidentified mechanisms.

The delayed activation of Rac1 and Cdc42 suggests that in spontaneous cell motility, that is, without active stimulation by growth factors, their primary role may not be in initiating protrusion, as is currently supposed¹⁵. Although the early, low activation level of Rac1 we observe may still be sufficient for this role, the peak activities for Rac1 and Cdc42 are localized closer to the sites of maturing adhesions ($\sim 1.8 \mu\text{m}$ behind the protruding leading edge¹⁶). Rac1 and Cdc42 seem to be affecting protrusion by regulating adhesion dynamics^{17–19}, probably by reinforcing adhesion sites that are formed to balance increasing protrusive forces at the leading edge^{20,21}. In addition, our data support the notion that Rac1 operates as an antagonist to RhoA¹². RhoA activity is rapidly suppressed as Rac1 reaches its maximum activation. RhoA is usually thought to be a mediator of contractility. Here, we propose that the synchronized activation of RhoA with protrusion relates to a second function of RhoA as an initiator of

actin polymerization at the onset of the protrusion–retraction cycle, possibly via its ability to activate members of the formin family such as mDia^{22,23}. In addition, the RhoA–mDia pathway is known to stabilize microtubules specifically in leading edge adhesions²⁴. This may contribute to the activation of Rac1 we observed at adhesion sites, perhaps via positive feedback between microtubule growth and Rac1 activation²⁵ and/or via the engagement of integrins that regulate the coupling of Rac1 to its effector Pak1²⁶, which is known to exhibit mutually positive feedback¹⁸.

We present two complementary approaches to *in situ* analysis of cellular pathways. Simultaneous imaging presents unprecedented spatial and temporal resolution to explore the relationship between two pathways, whereas 'computational multiplexing' enables correlation of many signalling activities. Computational multiplexing makes use of the spontaneous activation of pathways by random local changes in signalling molecules due to low level stochastic stimulation of receptors, variations in concentration, and so on. Thus, it is potentially less perturbing than methods that rely on acute cell stimulation to initiate a signalling cascade. The sensitivity required to capture the coupling of spontaneous molecular activities is achieved for two reasons: first, by very local image measurement, below the diffusion radius of signalling molecules, our analysis captures the immediate relationship between the states of a signalling pathway and its morphological outputs. Although the activity levels of pathway components and hence the pathway outputs may vary between cellular locations and between cells, the hierarchy and timing between pathway components are conserved. Second, measurement noise is greatly reduced by averaging thousands of local relationships between activities. Together, these properties permit the pathway to be reconstructed despite significant cell-to-cell heterogeneity.

We chose to illustrate computational multiplexing for a pathway controlling cell morphology, enabling us to use cell morphological dynamics as a reference to align distinct signalling events observed in different experiments. However, the method is readily scalable to more and diverse activities. Relationships between signals can also be inferred in pathways without spatial cues. In this case, one of the biosensors may itself be used as the common reference in a series of experiments where it is imaged simultaneously in pairings with other biosensors. Again, the statistical approaches described for computational multiplexing will be useful to integrate into a pathway model the heterogeneous and complex responses made visible by simultaneous imaging of different biosensor combinations.

METHODS SUMMARY

Biosensors. The three biosensors were made and used as originally described^{4–6}, with modifications where noted in the text. The Cdc42 biosensor was modified for multiplex imaging by removing EGFP and replacing it with an Alexa750 dye. This produced a biosensor with wavelengths orthogonal to those of the RhoA biosensor. The new Cdc42 biosensor was fused to maltose-binding protein to enhance solubility. For control studies a new dual-chain RhoA biosensor was constructed by eliminating the linker in the published RhoA biosensor (see Supplementary Information).

Live cell imaging. Biosensors were imaged as previously described^{27,28}. In order to obtain sufficient signal/noise for multiplexing, cells with biosensor were filmed at 10-s intervals, using a $40\times/1.3$ NA objective and binning, resulting in an effective pixel size in object space of 330 nm.

Ratiometric corrections. Raw images were aligned and noise-filtered to allow ratiometric correction of volume effects, bleed-through and variation in sensor concentration.

Image analysis. Corrected biosensor images were segmented and cell edge displacements tracked as in ref. 13. Sampling windows of $0.9 \mu\text{m}$ depth and 1.8 – $3 \mu\text{m}$ width were constructed to follow morphological changes at a fixed distance from the cell edge. For each window, biosensor activation time courses were recorded. For windows placed at the cell edge, a time course of protrusion/retraction velocity was recorded additionally. Coupling of two activity time courses was analysed per window by Pearson's cross-correlation function. Subsequently, for a cell the per-window correlation functions were averaged over all windows following the edge at a specific distance D (Fig. 1i). Per-cell correlation functions were averaged over multiple cells and statistically analysed

by bootstrap sampling to determine the significance and time lag of the coupling between two activities. All procedures are detailed in Supplementary Methods.

Received 5 January; accepted 24 June 2009.

Published online 19 August; corrected 3 September 2009 (see full-text HTML version for details).

1. Jaffe, A. B. & Hall, A. Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* **21**, 247–269 (2005).
2. Burridge, K. & Wennerberg, K. Rho and Rac take center stage. *Cell* **116**, 167–179 (2004).
3. Ridley, A. J. *et al.* Cell migration: integrating signals from front to back. *Science* **302**, 1704–1709 (2003).
4. Kraynov, V. *et al.* Localized Rac activation dynamics visualized in living cells. *Science* **290**, 333–337 (2000).
5. Nalbant, P., Hodgson, L., Kraynov, V., Touthkine, A. & Hahn, K. M. Activation of endogenous Cdc42 visualized in living cells. *Science* **305**, 1615–1619 (2004).
6. Pertz, O., Hodgson, L., Klemke, R. L. & Hahn, K. M. Spatiotemporal dynamics of RhoA activity in migrating cells. *Nature* **440**, 1069–1072 (2006).
7. Kurokawa, K. & Matsuda, M. Localized RhoA activation as a requirement for the induction of membrane ruffling. *Mol. Biol. Cell* **16**, 4294–4303 (2005).
8. Nobes, C. D. & Hall, A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* **81**, 53–62 (1995).
9. Rottner, K., Hall, A. & Small, J. V. Interplay between Rac and Rho in the control of substrate contact dynamics. *Curr. Biol.* **9**, 640–649 (1999).
10. Nimnual, A. S., Taylor, L. J. & Bar-Sagi, D. Redox-dependent downregulation of Rho by Rac. *Nature Cell Biol.* **5**, 236–241 (2003).
11. Arthur, W. T., Burridge, K. & Rho, A. Inactivation by p190RhoGAP regulates cell spreading and migration by promoting membrane protrusion and polarity. *Mol. Biol. Cell* **12**, 2711–2720 (2001).
12. Ohta, Y., Hartwig, J. H. & Stossel, T. P. FilGAP, a Rho- and ROCK-regulated GAP for Rac binds filamin A to control actin remodelling. *Nature Cell Biol.* **8**, 803–814 (2006).
13. Machacek, M. & Danuser, G. Morphodynamic profiling of protrusion phenotypes. *Biophys. J.* **90**, 1439–1452 (2006).
14. Marguet, D., Lenne, P. F., Rigneault, H. & He, H. T. Dynamics in the plasma membrane: how to combine fluidity and order. *EMBO J.* **25**, 3446–3457 (2006).
15. Raftopoulos, M. & Hall, A. Cell migration: Rho GTPases lead the way. *Dev. Biol.* **265**, 23–32 (2004).
16. Zaidel-Bar, R., Ballestrem, C., Kam, Z. & Geiger, B. Early molecular events in the assembly of matrix adhesions at the leading edge of migrating cells. *J. Cell Sci.* **116**, 4605–4613 (2003).
17. Del Pozo, M. A. *et al.* Integrins regulate GTP-Rac localized effector interactions through dissociation of Rho-GDI. *Nature Cell Biol.* **4**, 232–239 (2002).
18. Nayal, A. *et al.* Paxillin phosphorylation at ser273 localizes a GIT1-PIX-PAK complex and regulates adhesion and protrusion dynamics. *J. Cell Biol.* **173**, 587–599 (2006).
19. ten Klooster, J. P., Jaffer, Z. M., Chernoff, J. & Hordijk, P. L. Targeting and activation of Rac1 are mediated by the exchange factor β -Pix. *J. Cell Biol.* **172**, 759–769 (2006).
20. Ji, L., Lim, J. & Danuser, G. Fluctuations of intracellular forces during cell protrusion. *Nature Cell Biol.* **10**, 1393–1400 (2008).
21. Choi, C. K. *et al.* Actin and alpha-actinin orchestrate the assembly and maturation of nascent adhesions in a myosin II motor-independent manner. *Nature Cell Biol.* **10**, 1039–1050.
22. Narumiya, S., Ishizaki, T. & Watanabe, N. Rho effectors and reorganization of actin cytoskeleton. *FEBS Lett.* **410**, 68–72 (1997).
23. Yamana, N. *et al.* The Rho-mDia1 pathway regulates cell polarity and focal adhesion turnover in migrating cells through mobilizing Apc and c-Src. *Mol. Cell Biol.* **26**, 6844–6858 (2006).
24. Palazzo, A., Cook, T., Alberts, A. & Gundersen, G. mDia mediates Rho-regulated formation and orientation of stable microtubules. *Nature Cell Biol.* **3**, 723–729 (2001).
25. Rodriguez, O. C. *et al.* Conserved microtubule-actin interactions in cell movement and morphogenesis. *Nature Cell Biol.* **5**, 599–609 (2003).
26. Del Pozo, M. A., Price, L. S., Alderson, N. B., Ren, X. D. & Schwartz, M. A. Adhesion to the extracellular matrix regulates the coupling of the small GTPase Rac to its effector PAK. *EMBO J.* **19**, 2008–2014 (2000).
27. Hodgson, L., Nalbant, P., Shen, F. & Hahn, K. Imaging and photobleach correction of Mero-CBD, sensor of endogenous Cdc42 activation. *Methods Enzymol.* **406**, 140–156 (2006).
28. Hodgson, L., Shen, F. & Hahn, K. Biosensors for characterizing the dynamics of Rho family GTPases in living cells. *Curr. Protoc. Cell Biol.* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge funding from the Swiss National Science Foundation and the Novartis Foundation, formerly the Ciba-Geigy Jubilee Foundation (M.M.), NIH T32 GM008719 and NIH F30 HL094020 (C.W.), NIH R01 GM57464 (K.M.H.), NIH R01 GM71868 (G.D.), and the Cell Migration Consortium, grant U54 GM064346 from NIGMS (G.D. and K.M.H.).

Author Contributions M.M. initiated the project, conceptualized the idea of computational multiplexing, wrote all image analysis software pertinent to multiplexing, and contributed to the writing of the manuscript; L.H. developed simultaneous imaging of RhoA and Cdc42, including the modification and validation of the meroCBD probe, developed the intermolecular RhoA sensor, including controls and validation, studying the effects of biosensor stoichiometry and expression level, developed the new version of the Rac biosensor, and contributed to writing of the manuscript; C.W. produced stable cell lines of the intermolecular RhoA biosensor and conducted imaging experiments for the comparison of intra- and intermolecular biosensor designs; H.E. contributed simulations for validation of the correlation analysis and assisted with image processing; O.P. and P.N. contributed image data of RhoA and Cdc42 activity, respectively; A.A. and G.L.J. contributed valuable advice and unpublished reagents; K.M.H. and G.D. coordinated the study and wrote the final version of the manuscript and supplement.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.M.H. (khahn@med.unc.edu) or G.D. (gdanuser@scripps.edu).

LETTERS

A genetically encoded photoactivatable Rac controls the motility of living cells

Yi I. Wu^{1,3}, Daniel Frey⁴, Oana I. Lungu^{1,2,3}, Angelika Jaehrig^{1,3}, Ilme Schlichting⁴, Brian Kuhlman^{2,3} & Klaus M. Hahn^{1,3}

The precise spatio-temporal dynamics of protein activity are often critical in determining cell behaviour, yet for most proteins they remain poorly understood; it remains difficult to manipulate protein activity at precise times and places within living cells. Protein activity has been controlled by light, through protein derivatization with photocleavable moieties¹ or using photoreactive small-molecule ligands². However, this requires use of toxic ultraviolet wavelengths, activation is irreversible, and/or cell loading is accomplished via disruption of the cell membrane (for example, through microinjection). Here we have developed a new approach to produce genetically encoded photoactivatable derivatives of Rac1, a key GTPase regulating actin cytoskeletal dynamics in metazoan cells^{3,4}. Rac1 mutants were fused to the photoreactive LOV (light oxygen voltage) domain from phototropin^{5,6}, sterically blocking Rac1 interactions until irradiation unwound a helix linking LOV to Rac1. Photoactivatable Rac1 (PA-Rac1) could be reversibly and repeatedly activated using 458- or 473-nm light to generate precisely localized cell protrusions and ruffling. Localized Rac activation or inactivation was sufficient to produce cell motility and control the direction of cell movement. Myosin was involved in Rac control of directionality but not in Rac-induced protrusion, whereas PAK was required for Rac-induced protrusion. PA-Rac1 was used to elucidate Rac regulation of RhoA in cell motility. Rac and Rho coordinate cytoskeletal behaviours with seconds and submicrometre precision^{7,8}. Their mutual regulation remains controversial⁹, with data indicating that Rac inhibits and/or activates Rho^{10,11}. Rac was shown to inhibit RhoA in mouse embryonic fibroblasts, with inhibition modulated at protrusions and ruffles. A PA-Rac crystal structure and modelling revealed LOV–Rac interactions that will facilitate extension of this photoactivation approach to other proteins.

Recent NMR studies revealed the mechanism of a protein light switch in *Avena sativa* phototropin1 (refs 6, 12): a flavin-binding LOV2 domain interacts with a carboxy-terminal helical extension (J α) in the dark. Photon absorption leads to formation of a covalent bond between Cys 450 and the flavin chromophore, causing conformational changes that result in dissociation and unwinding of the J α helix. We fused the complete LOV2–J α sequence (404–547) to the amino terminus of a constitutively active Rac1, anticipating that the LOV domain in its closed conformation would block the binding of effectors to Rac1, and that light-induced unwinding of the J α helix would release steric inhibition, leading to Rac1 activation (Fig. 1a). Sampling of different junctional sequences in pull-down assays revealed that connecting Leu 546 of LOV2–J α to Ile 4 of Rac1 led to substantial reduction in Rac1 binding to its effector PAK (Fig. 1b and Supplementary Fig. 1a). To ensure that the photoactivatable Rac1 would induce no dominant-negative effects and that its activity would not be subject to upstream regulation, mutations were introduced to

abolish GTP hydrolysis and diminish interactions with nucleotide exchange factors, guanine nucleotide dissociation inhibitors (Q61L) and GTPase activating proteins (E91H and N92H) (Supplementary Fig. 2 and Supplementary text ‘Characterization of Rac1 constructs’). This resulted in the photoactivatable analogue of Rac1 (PA-Rac1) used in the following studies. Pull-down assays showed that PA-Rac1 has greatly reduced affinity for its effector protein PAK in the dark, as does a PA-Rac1 construct containing a light-insensitive LOV2 mutation (C450A)¹³. Effector binding was restored in a PA-Rac1 construct containing a LOV2 mutant (I539E)¹⁴ which mimics the unfolded ‘lit state’ (Fig. 1b and Supplementary Fig. 1b). Isothermal titration experiments indicated that the dark and lit state mutants of PA-Rac1 differed tenfold in effector binding (200 nM versus 2 μ M) (Supplementary Fig. 3 and Supplementary Table 1), with lit state effector affinity similar to that of native Rac¹⁵.

Activation of PA-Rac1 was examined in HeLa cells expressing a YFP fusion of PA-Rac1 to gauge expression level. The cells remained quiescent when illuminated with wavelengths longer than flavin absorbance (515, 568 or 633 nm, data not shown), but within seconds after switching to 458 nm, lamellipodial protrusions and membrane ruffles appeared around the cell edges (Fig. 1c and Supplementary Movie 1). To show that this effect was due to PA-Rac1, kymographs were used to quantify maximum protrusion length; irradiation of PA-Rac1 elicited protrusions that were four times as long as those seen in cells expressing either LOV domain alone or the light-insensitive PA-Rac1(C450A) mutant (Supplementary Fig. 4). An important advantage of PA-Rac1 is its ability to control precisely the subcellular location of Rac activation. We first examined this in mouse embryo fibroblasts (MEFs) stably expressing PA-Rac1, and cultured without serum to minimize cell activity before irradiation. Irradiation of 20- μ m spots at the cell edge generated large protrusions clearly localized next to the point of irradiation (Fig. 1d and Supplementary Movie 2). Repeated irradiation led first to ruffles and then to protrusion. YFP–actin, YFP–PAK and YFP–Arp3 revealed actin polymerization at the edge of the Rac-induced protrusions with associated translocation of downstream effectors, and induction of localized PAK phosphorylation was shown by immunostaining (Supplementary Figs 5 and 6 and Supplementary Movies 3 and 4). Movement of a laser spot to different positions led to cessation of ruffling or protrusion at the initial irradiation position and new activities appearing where the laser spot was brought to rest (HeLa cells, Supplementary Movie 5), demonstrating reversible activation. In MEF cells, more prone to movement than HeLa cells, complex shape changes were produced by ‘painting’ the cell with the laser spot (Supplementary Movie 6). The area of protrusions in MEF cells was dependent on light dosage, indicating the valuable ability to control the level of Rac1 activation (Supplementary Fig. 7). PA-Rac1 diffusion was analysed using fluorescence recovery after photobleaching (FRAP) and using PA-Rac1 tagged with photoactivatable GFP¹⁶ (Supplementary

¹Department of Pharmacology, ²Department of Biochemistry and Biophysics, and ³Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁴Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahn-Strasse 29, 69120 Heidelberg, Germany.

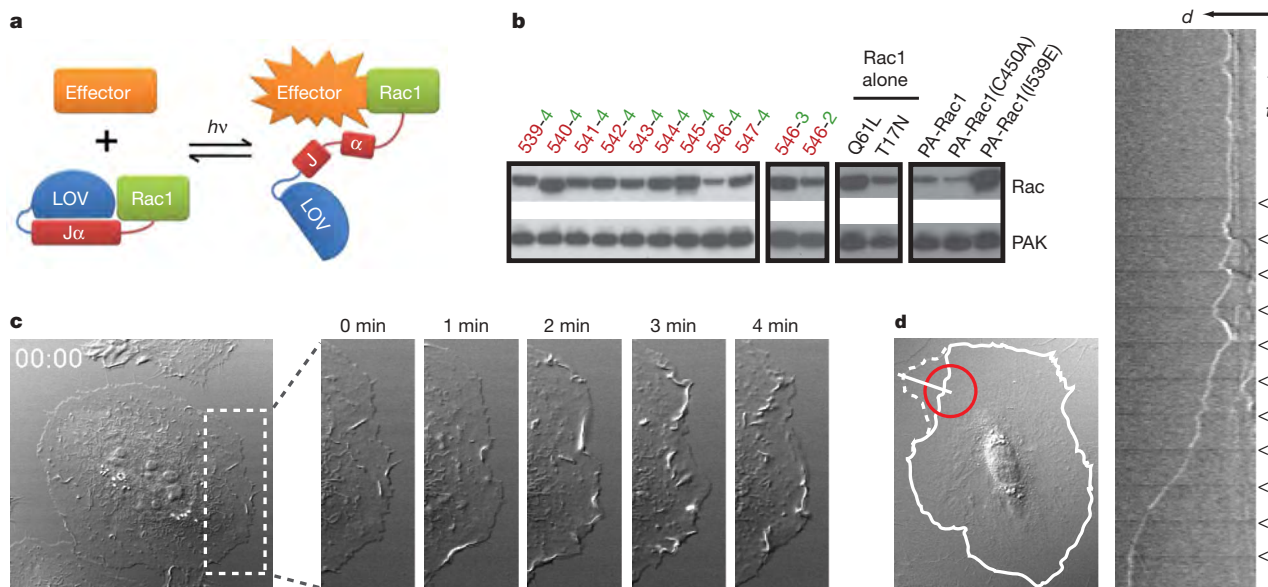


Figure 1 | Engineering and *in vivo* characterization of a photoactivatable Rac1 (PA-Rac1). **a**, Cartoon representation of PA-Rac1 design. *hν*, irradiation. **b**, Pull down of PA-Rac1 constructs with PAK in the dark. Truncations of LOV and Rac at their linkage point were tested: 539–547 in red, terminal amino acid of Jα; 2–4 in green, first residue of Rac1. 546–4 showed the strongest inhibition. PA-Rac1, 546–4 Q61L/E91H/N92H; PA-Rac1(C450A), light-insensitive mutant; PA-Rac1(I539E), lit state mutant. Pull down by constitutively active (Q61L) and dominant negative (T17N) mutants are included for comparison with PA-Rac1. **c**, Whole-cell

irradiation of a HeLa cell expressing PA-Rac1 (minutes after irradiation, DIC, short axis of box = 20 μm). **d**, Spatial control of Rac1 activity. A 20-μm circle (red) was irradiated every 60 s in serum-starved MEF cells. Solid line, cell border at time 0; dotted line, 10 min after initial light pulse. Little movement of the cell border was detected, except adjacent to the point of irradiation. The kymograph (taken using white line, 20 μm), shows the initial formation of ruffles after each pulse, followed by protrusion (arrowheads indicate irradiation pulses).

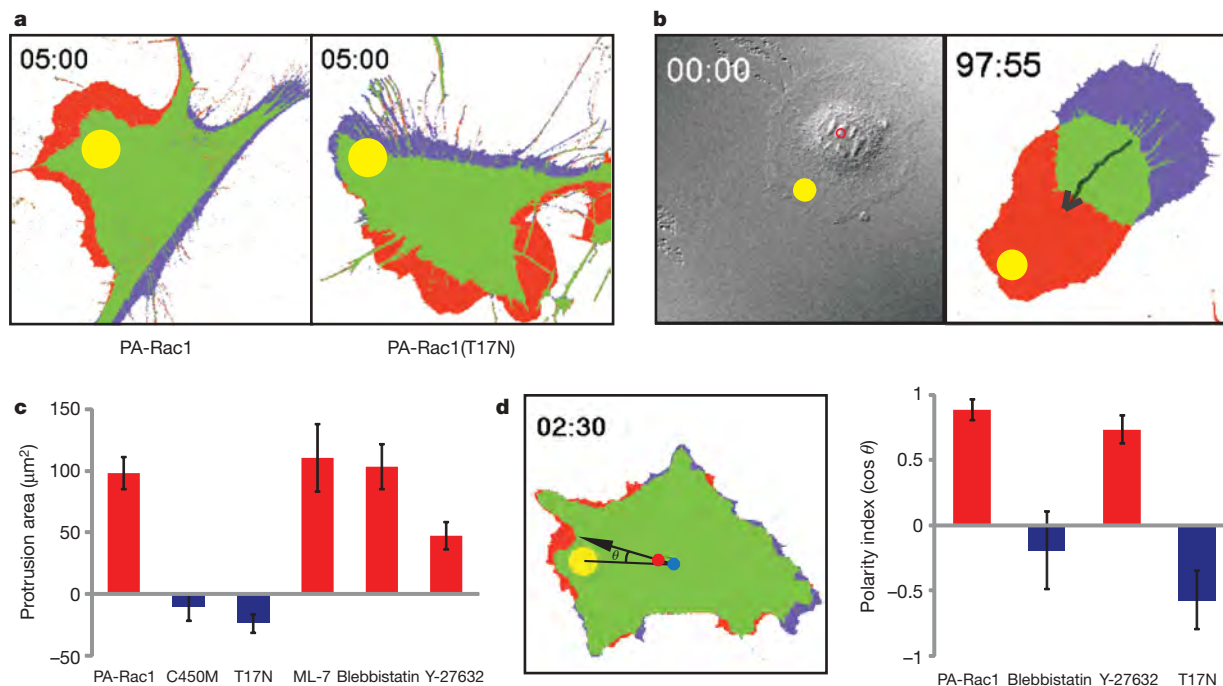


Figure 2 | Localized activation or inactivation of PA-Rac1 induces myosin-dependent migration. **a**, Protrusion/retraction map after a single pulse of activating illumination. MEFs expressing PA-Rac1 (left) generated protrusions at the site of irradiation (red) and retraction at the opposite side of the cell (blue) (in all 50 cells studied). Irradiation of the dominant-negative T17N mutant of PA-Rac1 (right) produced retraction near the point of irradiation, with protrusion in area(s) other than the site of irradiation (in all 25 cells studied). **b**, Repeated activation of PA-Rac1 at the cell edge induces directional migration. (MEF, 2-min intervals, average 0.8 μm movement per pulse, *n* = 6.) **c**, Localized activation of PA-Rac1 in the

presence of ML-7 (MLCK inhibitor, 1 μM), blebbistatin (myosin II ATPase inhibitor, 1 μM) or Y-27632 (ROCK inhibitor, 10 μM). Protrusions analysed as in panel **a**. **d**, Effect of myosin or ROCK inhibition on the ability of Rac1 to specify the direction of movement. The cosine of the angle between two lines (from the irradiation spot to the cell centroid at time 0, from the centroid at time 0 to the centroid at the end of the experiment) indicated how much the cell deviates from the direction specified by local irradiation. For **c**, **d**, *n* > 25; means ± 95% confidence intervals; throughout Fig. 3 irradiation at 458 nm, spot diameter = 10 μm; time shown is in minutes and seconds.

Fig. 8 and Supplementary Movies 7 and 8), indicating that PA-Rac1 diffuses more slowly than cytosolic proteins, probably because it is membrane bound ($10\text{ }\mu\text{m}$ spot, FRAP $D = 0.55\text{ }\mu\text{m}^2\text{ s}^{-1}$ or $t_{1/2} = 12.1\text{ s}$; PA-GFP $t_{1/2} = 14.6\text{ s}$). The half-life of dark recovery for PA-Rac1 was determined to be 43 s at room temperature. Simulation using this value indicated that, for two adjacent $10\text{-}\mu\text{m}$ spots, the unirradiated spot will achieve at most 7.5% the activation of the irradiated region (Supplementary Fig. 8c). Together these studies validate PA-Rac1 as a robust, genetically encoded and reversible caged protein effective in living cells.

We used PA-Rac1 to ask whether localized Rac activation is sufficient to specify cell polarity. In MEF cells, activating Rac1 at one spot near the cell edge not only generated protrusion locally, but also produced retraction on the opposite side of the cell (Fig. 2a and Supplementary Movie 9). To test whether this cross-cell coordination was due to a gradient of Rac1 activity, we fused the LOV domain to a dominant-negative mutant of Rac1 using the same linkage as in PA-Rac1. Irradiation of this PA-Rac1(T17N) led to nearby retraction rather than protrusion, and now generated protrusion in other areas of the cell (Fig. 2a and Supplementary Movies 10 and 11). The ability of Rac1 alone to control polarized movement was confirmed by repeated irradiation at the cell edge, which could be used to produce prolonged cell movement by generating consistent coordinated extension and retraction (MEF cells, Fig. 2b and Supplementary Movie 12; HEK293 cells, Supplementary Movie 13). In contrast to MEF and HEK293 cells, HeLa cells showed localized protrusion but

could not be induced to retract or move simply by activating Rac (Supplementary Movie 3), indicating that Rac-induced motility is subject to modulation by other pathways.

PA-Rac1 enabled control of Rac1 activity without the prior cellular compensation seen with other techniques, that is, mutation or altered expression. Using this advantage, we examined the role of myosin, a key mediator of actin-based contractility, in Rac-induced motility. Global inhibition of myosin activity using the myosin ATPase inhibitor blebbistatin or the myosin light chain kinase inhibitor ML-7 strongly affected Rac's ability to specify the direction of cell movement, but minimally affected Rac-induced protrusion (Fig. 2c, d). Myosin may mediate Rac's control of directionality through induction of tail retraction¹⁷, contraction of the cell cortex to direct protrusive force¹⁸, or coupling of actin to adhesions differently at the front and rear¹⁹. In contrast, inhibition of PAK was found to strongly affect Rac-induced protrusion (Supplementary Fig. 9). Inhibition of the Rho-activated kinase ROCK using Y27632 suggested a role for ROCK in Rac-induced protrusion, but these results must be interpreted with caution owing to known off-target effects²⁰.

Where and how Rac regulates Rho *in vivo* remains largely unknown; this was examined by using PA-Rac1 together with a RhoA biosensor⁸. Localized activation of Rac1 led to immediate inhibition of RhoA, and this inhibition spread outwards from the irradiated spot (Fig. 3a and Supplementary Movie 14). This was not simply an artefact of biosensor photobleaching, as irradiating the photoinactive C450M mutant (Fig. 2c) of PA-Rac1 led to localized

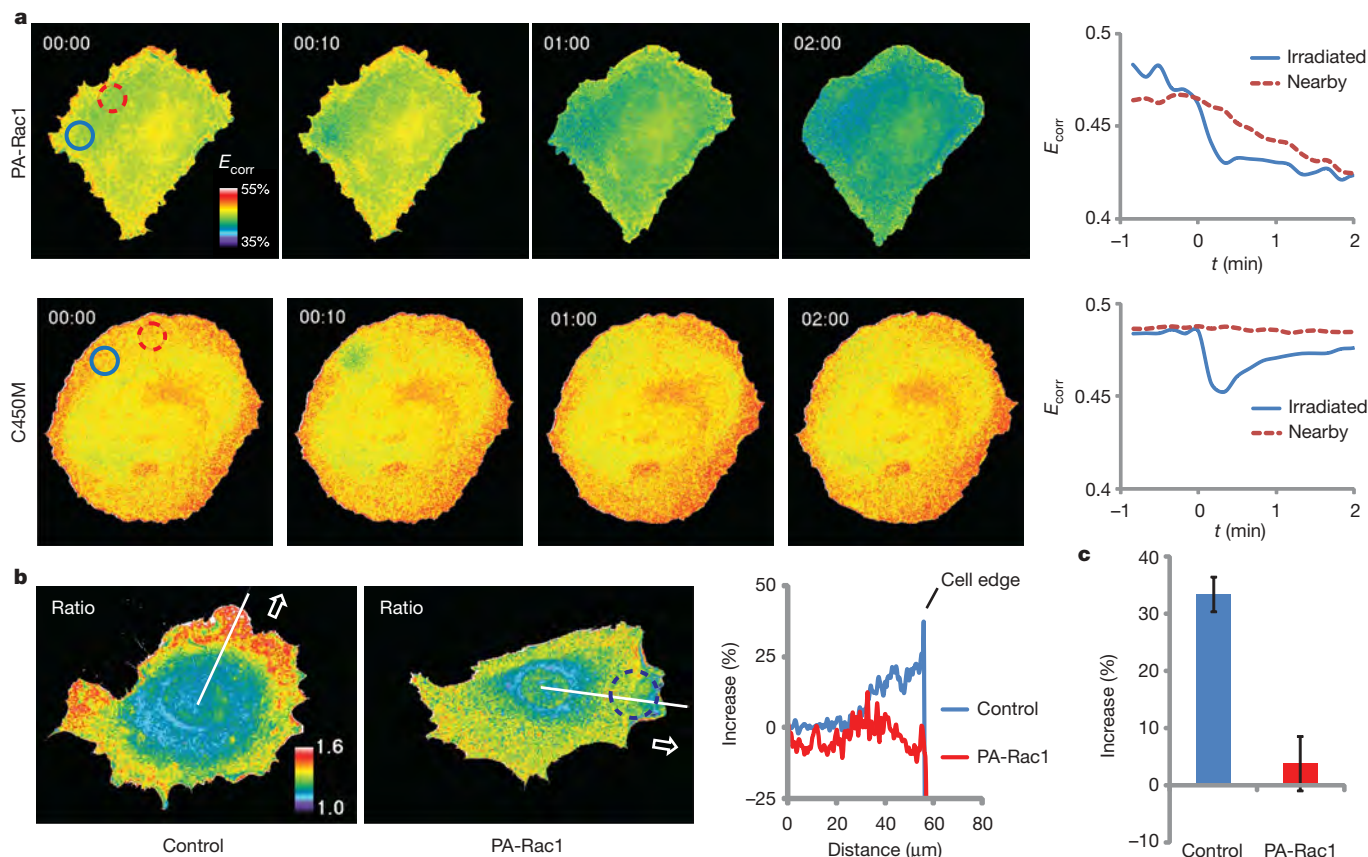


Figure 3 | Inhibition of RhoA by PA-Rac1. **a**, HeLa cells expressing RhoA biosensor and either PA-Rac1 or its C450M photoinactive mutant, illuminated in a $10\text{-}\mu\text{m}$ circle with a single pulse of 473-nm light. Changes in the FRET efficiency (E_{corr}) of the RhoA biosensor, indicative of RhoA activation, are shown in pseudocolour and as plots of average FRET efficiency within the irradiated circle (blue) and a nearby circle (red). In the PA-Rac1 cells, the irradiated spot showed bleaching of the biosensor followed by a relatively constant level of reduced RhoA activity. The nearby spot showed no bleaching, but a gradual decrease in RhoA activity reaching

the low level achieved in the irradiated spot ($n = 3$ cells). In the control cells (C450M), the biosensor returned to near initial activation readouts after bleaching, and no change was seen in the nearby spot ($n = 3$ cells). Time shown is in minutes and seconds. **b**, RhoA activation in constitutive pseudopods^{8,21} versus pseudopods induced by PA-Rac1 (473 nm , $20\text{-}\mu\text{m}$ circle shown, Supplementary Movie 15). **c**, The bar graph shows the per cent increase in biosensor FRET/CFP ratio in the region $1\text{ }\mu\text{m}$ from the cell edge versus the mean of the flat region at the left of the line scan. Means \pm 95% confidence intervals, 18 lines from 6 cells per bar.

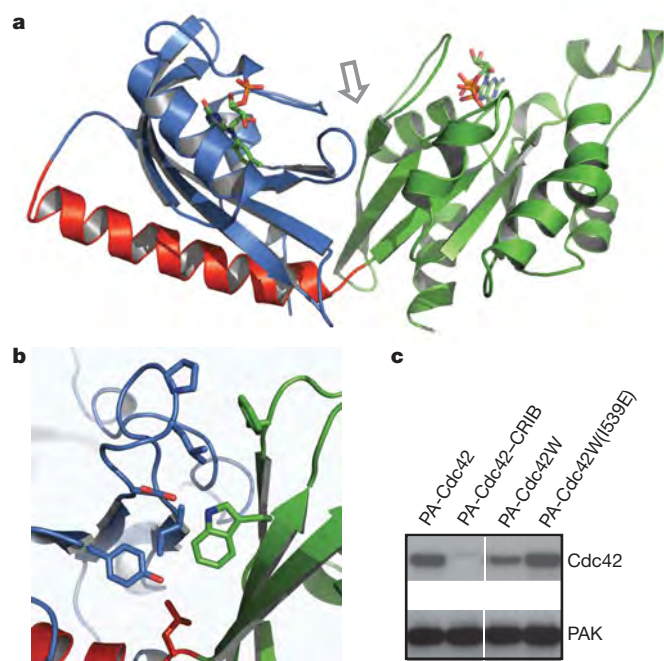


Figure 4 | Crystallization and structural modelling of PA-Rac1. **a**, Dark state crystal structure of PA-Rac1. Blue, LOV domain; red, J α helix; green, Rac1. **b**, Interacting residues at the LOV–Rac interface (arrow in panel **a**), including Trp 56. **c**, Mutating Cdc42 to include the Trp involved in stabilizing the LOV2–Rac1 interaction substantially improved LOV inhibition of Cdc42. Lane 1, PA-Cdc42; linking LOV to Cdc42 using the same truncations that produced good inhibition for Rac does not inhibit Cdc42–PAK binding. Lane 2, PA-Cdc42–CRIB; covalently linking the CRIB domain of PAK to PA-Cdc42 blocks PAK binding. Lane 3, PA-Cdc42(F56W) (PA-Cdc42W); introduction of the tryptophan substantially improves LOV inhibition of Cdc42 binding to PAK. Lane 4, lit state mutant of PA-Cdc42(F56W) (PA-Cdc42W(I539E)), showing that Cdc42 inhibition is sensitive to the lit/dark state of the LOV domain. Supplementary Movie 16 and Supplementary Fig. 14 demonstrate the ability of PA-Cdc42(F56W) to produce filopodia and protrusions in living cells.

biosensor photobleaching and recovery, but no prolonged local inhibition or wave of inhibition (Fig. 3a). There were marked differences between constitutive MEF protrusions and protrusions induced by pulsed PA-Rac1 irradiation. In contrast to constitutive protrusions, RhoA activity was greatly reduced in protrusions induced by PA-Rac (Fig. 3b). Inhibition of RhoA seems to be compartmentalized or controlled kinetically when Rac is activated in the context of normal motility, as both active Rac and active Rho are seen at the leading edge^{7,8,21}. PA-Rac activation led to large ruffles moving from the site of irradiation rearwards towards the nucleus (Supplementary Movie 15), suggesting that Rac regulates rearward membrane flow. In control experiments, irradiation of cells expressing the photoinactive C450M mutant did not produce polarized ruffling or show reduced RhoA activity (data not shown).

To understand the structural basis of the PA-Rac1 switch for future application to other proteins, we performed Rosetta structure prediction simulations²² on several LOV2–Rac1 constructs, and determined high-resolution crystal structures of photoactive and inactive PA-Rac1 in the dark state. The crystal structures confirmed that the LOV domain occludes effector binding in the dark state (Fig. 4a and Supplementary Table 2). LOV-J α adopted a closed conformation that superimposes with the recently published structure of isolated LOV-J α (ref. 23). In the conformational ensemble predicted by simulations of the dark state, the effector binding site of Rac was sterically blocked by the LOV domain in a majority of the low-energy models (Supplementary Tables 3–5 and Supplementary Figs 10–13). Consistent with pull-down assays (Fig. 1b and Supplementary Fig. 1a), adding or removing even one residue from the connection

between LOV and Rac resulted in conformational ensembles with exposed effector binding sites. In the dark state, Rac was seen to form an extensive interface with the LOV domain (Fig. 4b), occluding Rac binding interactions. Given the substantial structural similarity between Rac1 and Cdc42, we hypothesized that the LOV domain could also be used to cage Cdc42. However, the linkage used for PA-Rac1 failed to reduce Cdc42 binding to PAK (Fig. 4c and Supplementary Fig. 1d). Using the PA-Rac1 crystal structure as a template, a model was built of the Cdc42–LOV dark state. At the interface between Rac and LOV a hydrophobic cluster is formed between residues Phe37 and Trp 56 from Rac and Leu 422, Pro 423, Ile 428, Tyr 508 and Leu 546 from LOV. Consistent with this being a weak, non-evolved interaction, most of the hydrogen-bonding potential at the Rac–LOV interface is satisfied by buried and partially buried water molecules instead of inter-domain hydrogen bonds (Fig. 4b). This interface model was used to identify a mutation to Cdc42, F56W, at the Rac–LOV interface, that was predicted to stabilize the dark state. Pull-down assays showed that this mutation substantially improves dark-state inhibition of PAK binding, and produces differential affinity for Cdc42 effector in the dark versus the lit state (Fig. 4d). In living cells, irradiation of the mutated PA-Cdc42 led to production of filopodia and in some cases protrusions and/or ruffles, consistent with Cdc42 induction of filopodia and activation of Rac²⁴ (Supplementary Fig. 14 and Supplementary Movie 16). These results argue that PA-Rac1 can serve as a blueprint for engineering other caged GTPases.

We have engineered genetically encoded photoactivatable Rac1 analogues that enable precise spatial and temporal control of Rac activity in live cells, with reversible activation at 458 or 473 nm. Localized Rac activation or deactivation was sufficient to generate polarized cell movement. Rac could be activated without cellular compensation, enabling us to probe the role of myosin and PAK in Rac-mediated motility. Spatially regulated Rac inhibition of Rho was demonstrated in living cells. Structural studies indicate that a non-evolved interaction at the Rac–LOV interface can be engineered to cage other GTPases. This study and other recent work^{25–28} show that coupling genetically encoded light-modulated domains to other proteins provides a versatile new route to control protein activities in living cells.

METHODS SUMMARY

Imaging experiments were conducted on an Olympus FluoView 1000 laser scanning confocal microscope and an Olympus IX81-ZDC inverted microscope. Biosensor imaging was performed as previously described^{8,29}. Simultaneous biosensor imaging and activation of PA-Rac was achieved using a MAG Biosystems FRAP-3D add-on (Photometrics) for galvanometer control of laser position.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 March; accepted 30 June 2009.

Published online 19 August 2009.

- Goeldner, M. & Givens, R. *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules* (Wiley-VCH, 2005).
- Fortin, D. L. *et al.* Photochemical control of endogenous ion channels and cellular excitability. *Nature Methods* **5**, 331–338 (2008).
- Raftopoulos, M. & Hall, A. Cell migration: Rho GTPases lead the way. *Dev. Biol.* **265**, 23–32 (2004).
- Ridley, A. J. *et al.* Cell migration: integrating signals from front to back. *Science* **302**, 1704–1709 (2003).
- Christie, J. M., Salomon, M., Nozue, K., Wada, M. & Briggs, W. R. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. *Proc. Natl Acad. Sci. USA* **96**, 8779–8783 (1999).
- Harper, S. M., Neil, L. C. & Gardner, K. H. Structural basis of a phototropin light switch. *Science* **301**, 1541–1544 (2003).
- Kraynov, V. S. *et al.* Localized Rac activation dynamics visualized in living cells. *Science* **290**, 333–337 (2000).
- Pertz, O., Hodgson, L., Klemke, R. L. & Hahn, K. M. Spatiotemporal dynamics of RhoA activity in migrating cells. *Nature* **440**, 1069–1072 (2006).
- Burridge, K. & Wennerberg, K. Rho and Rac take center stage. *Cell* **116**, 167–179 (2004).

10. Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–410 (1992).
11. Sander, E. E., ten Klooster, J. P., van Delft, S., van der Kammen, R. A. & Collard, J. G. Rac downregulates Rho activity: reciprocal balance between both GTPases determines cellular morphology and migratory behavior. *J. Cell Biol.* **147**, 1009–1022 (1999).
12. Yao, X., Rosen, M. K. & Gardner, K. H. Estimation of the available free energy in a LOV2-J α photoswitch. *Nature Chem. Biol.* **4**, 491–497 (2008).
13. Salomon, M., Christie, J. M., Knieb, E., Lempert, U. & Briggs, W. R. Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. *Biochemistry* **39**, 9401–9410 (2000).
14. Harper, S. M., Christie, J. M. & Gardner, K. H. Disruption of the LOV-J α helix interaction activates phototropin kinase activity. *Biochemistry* **43**, 16184–16192 (2004).
15. Thompson, G., Owen, D., Chalk, P. A. & Lowe, P. N. Delineation of the Cdc42/Rac-binding domain of p21-activated kinase. *Biochemistry* **37**, 7885–7891 (1998).
16. Patterson, G. H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873–1877 (2002).
17. Vicente-Manzanares, M., Zareno, J., Whitmore, L., Choi, C. K. & Horwitz, A. F. Regulation of protrusion, adhesion dynamics, and polarity by myosins IIA and IIB in migrating cells. *J. Cell Biol.* **176**, 573 (2007).
18. Burridge, K. & Chrzanowska-Wodnicka, M. Focal adhesions, contractility, and signaling. *Annu. Rev. Cell Dev. Biol.* **12**, 463–518 (1996).
19. Giannone, G. et al. Lamellipodial actin mechanically links myosin activity with adhesion-site formation. *Cell* **128**, 561–575 (2007).
20. Davies, S. P., Reddy, H., Caivano, M. & Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.* **351**, 95–105 (2000).
21. Kurokawa, K. & Matsuda, M. Localized RhoA activation as a requirement for the induction of membrane ruffling. *Mol. Biol. Cell* **16**, 4294–4303 (2005).
22. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
23. Halavaty, A. S. & Moffat, K. N- and C-terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor phototropin 1 from *Avena sativa*. *Biochemistry* **46**, 14001–14009 (2007).
24. Nobes, C. D. & Hall, A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* **81**, 53–62 (1995).
25. Lee, J. et al. Surface sites for engineering allosteric control in proteins. *Science* **322**, 438–442 (2008).
26. Möglich, A., Ayers, R. A. & Moffat, K. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *J. Mol. Biol.* **385**, 1433–1444 (2008).
27. Strickland, D., Moffat, K. & Sosnick, T. R. Light-activated DNA binding in a designed allosteric protein. *Proc. Natl Acad. Sci. USA* **105**, 10709–10714 (2008).
28. Leung, D. W., Otomo, C., Chory, J. & Rosen, M. K. Genetically encoded photoswitching of actin assembly through the Cdc42-WASP-Arp2/3 complex pathway. *Proc. Natl Acad. Sci. USA* **105**, 12797–12802 (2008).
29. Hodgson, L., Shen, F. & Hahn, K. M. Biosensors for characterizing the dynamics of Rho family GTPases in living cells. *Curr. Protoc. Cell Biol.* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for help and constructs from W. Briggs, K. Moffat, A. Tripathy, G. Bokoch and K. Jacobson. Diffraction data were collected at the Swiss Light Source, beamline X10SA, Paul Scherrer Institute, Villigen, Switzerland. We thank the Dortmund-Heidelberg team for data collection, and A. Pauluhn and M. Fuchs for their support in setting up the beamline. This research was supported by the American Heart Association (Y.I.W.) and the National Institutes of Health (K.M.H. grants GM057464 and GM64346).

Author Contributions Y.I.W. initiated the project, demonstrated the validity of caging Rac with LOV, and performed the studies of Rac biological function. He was assisted by A.J. in cloning and protein expression. D.F. and I.S. determined and analysed the crystal structures. O.I.L. performed molecular modelling and isothermal calorimetry studies under the direction of B.K. K.M.H. coordinated the study and wrote the final version of the manuscript, based on contributions from all authors.

Author Information The structural coordinates of PA-Rac1 and its mutants have been submitted to the Protein Data Bank under accessions 2wkp (wild type), 2wkq (C450A) and 2wkr (C450M). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.M.H. (khahn@med.unc.edu) or Y.I.W. (yiwu@med.unc.edu).

METHODS

DNA cloning. The cDNA encoding the LOV2 domain of *Avena sativa* (oat) Phototropin1 (404–546), including the C-terminal helical extension (J α), was a gift from K. Moffat. Chimaeric fusion constructs consisting of LOV2–J α fused to Rac1 or Cdc42 were generated using an overlapping PCR approach so that precise junctional sequences could be engineered without being limited by restriction sites. These included truncations/extensions of the LOV2–J α C terminus (539–547), the N terminus (2–4) of the GTPases, or insertions of designed Schellman caps (KEAGADQI and KELKEAGADQI)¹. The QuickChange (Stratagene) protocol was used to introduce additional point mutations, including C450A or C450M, and I539E to mimic the dark and lit state of the LOV domain, respectively. PA-Rac1 was constructed as follows: LOV2–J α (404–546)–Rac1(4–192)/Q61L/E91H/N92H. These constructs were inserted into a pTriEx (Novagen) vector for transient expression in mammalian cells as well as in bacteria. For crystallization, C-terminally truncated PA-Rac1 (Δ 181–192) was subcloned into the pQE-30 vector (Qiagen). Fluorescent proteins mVenus², mCherry³ and mPA-GFP⁴ were inserted at the N terminus of the LOV domain with a short GSLS linker to monitor expression and subcellular localization. After initial characterization, PA-Rac1 with different fluorescent protein tags was subcloned into pBabe-TetCMV vector for retroviral production and establishment of stable MEF Tet-Off cell lines. High-fidelity Pfu Turbo DNA polymerase (Stratagene) was used in PCR reactions and all plasmids were verified by DNA sequencing.

Pull-down assay of effector binding. mVenus-tagged LOV2–J α and Rac1 fusion constructs were coexpressed with Flag-tagged PAK1 in HEK293 (LinXE) cells by transient transfection using Eugene 6 (Roche). The cells were lysed in 50 mM Tris pH 7.5, 150 mM NaCl and 1% Triton X-100 (lysis buffer) with addition of EDTA-free protease inhibitor cocktail (Roche). After brief centrifugation, the supernatants were incubated with Flag/M2-agarose (Sigma) followed by washes with lysis buffer, and elution with lysis buffer containing 200 μ g ml^{−1} 3 \times Flag peptide (Sigma). All procedures were done at 4 °C under red light, facilitated using Handee spin columns (Pierce). The purified protein complexes as well as cell lysates were fractionated on 4–12% NuPAGE gels (Invitrogen) followed by western blot analysis using antibodies against fluorescent protein (JL-8, Clontech) and PAK (N-20, Santa Cruz). Myc pull-down experiments were performed similarly using anti-Myc agarose and Myc peptide (Sigma) to investigate the interaction of PA–Rac constructs with p50RhoGAP.

Expression, purification and characterization of proteins used for crystallization. C-terminal truncated PA-Rac1 (Δ 181–192) was expressed in *Escherichia coli* strain XL-10 Gold (Stratagene) at 30 °C overnight. All purification steps were done under yellow light at 4 °C. Cells were lysed in 20 mM Tris pH 8.5, 50 mM NaCl, 5 mM MgCl₂ and 2 mM 2-mercaptoethanol. Protein was purified with a Ni-NTA-FastFlow column (Qiagen) exploiting the N-terminal 6 \times His tag. The elution was dialysed against 10 mM Tris pH 8.5, 20 mM NaCl, 5 mM MgCl₂ and 2 mM 2-mercaptoethanol. The protein was bound to a MonoQ column (GE Healthcare) and eluted with a linear gradient (0–250 mM NaCl in 50 CV). Fractions containing the protein were concentrated (30 kDa cutoff, Millipore) and further purified by Superose 6 gel filtration chromatography (GE Healthcare, 10 mM Tris pH 8.5, 20 mM NaCl, 5 mM MgCl₂ and 2 mM DTE). Before crystallization the protein was concentrated to 10 mg ml^{−1} (30 kDa cut-off, Millipore). All proteins were characterized spectroscopically. PA-Rac1 showed reversible light/dark conversion. The dark recovery rate of PA-Rac1 was measured as described previously⁵. The inactive C450A mutation showed no effect on light illumination whereas truncated C450M (Δ 181–192) was prone to aggregation on light illumination and, therefore, no data for light/dark conversion could be measured.

Crystallization. Crystallization was carried out under dimmed red light at 20 °C. PA-Rac1 and its C450A and C450M mutants were crystallized using the vapour diffusion method by mixing equal volumes of protein (10 mg ml^{−1}) and precipitant solution (100 mM calcium acetate/100 mM sodium cacodylate/12% (w/v) PEG 8000 or 4% (w/v) PEG 4000/100 mM potassium chloride). Yellow pencil-shaped crystals appeared overnight and grew to a final size of 50 \times 50 \times 1,000 μ m³ in a week. To avoid photoactivation, crystal handling was done by shielding the microscope light bulb with a 2-mm-thick RG630 filter (ITOS). Before cooling the crystals in liquid nitrogen they were transferred stepwise to precipitant solution supplemented with 20% (v/v) ethylene glycol for cryoprotection.

Data collection and structure determination. Diffraction data were collected at the X10SA beamline (Swiss Light Source) under standard cryogenic settings. Data were reduced using the XDS suite⁶ (Supplementary Table 2) and the structure was solved by molecular replacement⁷ using subsequently 2V0U⁸ and 1MH1⁹ as the initial models. During several rounds of refinement with PHENIX¹⁰ and manual model building in COOT¹¹, FMN, GTP, Mg²⁺, Ca²⁺

and solvent molecules were included in the model. Structures were validated using MOLPROBITY¹² and PROCHECK¹³ (see Supplementary Table 2 for final statistics).

Structural modelling for linker optimization. The Rosetta program^{14,15} was used to predict the dark state structure of LOV2–Rac1 based on the solved crystal structures of dark state LOV2⁸ (Protein Data Bank (PDB) code 2V0U) and Rac1⁹ (PDB code 1MH1). Structure prediction simulations were performed on LOV2–Rac1 545–4, 546–4 and 547–4 constructs. In these simulations, the torsion angles of the residues connecting the two proteins were optimized with Monte Carlo sampling. Using the Rosetta domain assembly protocol¹⁶, we first applied 1,000 Φ and Ψ backbone torsion angle movements of up to 180° each to three residues connecting LOV2 to Rac1 in a low-resolution representation. Small backbone torsion angle moves of up to 4° were then performed on a high-resolution representation of LOV2–Rac1, followed by a global repacking of all side-chain rotamers. After every 15 cycles of small moves and repacking, further repacking was restricted to the rotamers at the interface and next to the LOV2–Rac1 linkers. This sequence of refinement was repeated for a total of 150 cycles. Next, we adopted a series of small moves, global rotamer repacking, as well as backbone minimization within 5 residues of the LOV2–Rac1 linker for high-resolution optimization cycles. After every ten cycles, only rotamers at the interface and next to the LOV2–Rac1 linkers were repacked. A total of 100 such high resolution optimizations were used to generate models, which were further scored using Rosetta's energy function. One-thousand models, each representing a different folding trajectory, were generated per construct from simulations using the domain assembly protocol.

The complex structure of Rac3 and the CRIB domain of PAK4 (PDB code 2OV2) was used to model the interaction of CRIB-containing effectors with LOV2–Rac1 constructs. The crystal structure of Rac1⁹ (PDB 1MH1) was superimposed onto the complex structure by mapping the C α atoms of Rac1 onto those of Rac3. This derived complex structure was then superimposed onto the LOV2–Rac1 models to create model–CRIB complexes. Side-chain rotamers at the interface of each complex were optimized using rotamer repacking¹⁷. These complexes were scored using the Rosetta energy function. A low-scoring model–CRIB complex indicated the model could bind CRIB, whereas a high-scoring model–CRIB complex indicated clashes between atoms of the model and the CRIB domain, resulting in reduced binding.

Models generated in a simulation were grouped into clusters according to their pairwise root mean square deviation (r.m.s.d.). The r.m.s.d. (in Å) of the C α atom positions of each model from all other models in the simulation was calculated. Those models falling within a radius of 3 Å r.m.s.d. from each other were grouped into a cluster. A cluster member representing the centre of each cluster was chosen.

Isothermal titration calorimetry. Dark and lit state mimetics of PA-Rac1, C450A and I539E were cloned into a pTriEx vector with an N-terminal six histidine tag. Residues 65–150 of PAK1, comprising the extended CRIB domain, were cloned into a pET23 vector, with a C-terminal 6 \times histidine tag. All proteins were expressed in *E. coli* strain BL21(DE3) cells (Stratagene) at 16 °C overnight in the dark. Cells were lysed in 50 mM sodium phosphate pH 7.0, 300 mM NaCl and 5 mM MgCl₂. Proteins were purified under yellow light using TALON Metal Affinity Resin (Clontech) and eluted with 150 mM imidazole at pH 7.0. The proteins were dialysed against 50 mM sodium phosphate, 150 mM NaCl, 7.15 mM 2-mercaptoethanol, 5 mM MgCl₂ and 1% glycerol. Protein concentrations were quantified based on the estimated molar extinction coefficients (280 nm) of the corresponding polypeptides or the reported molar extinction coefficient (447 nm) of LOV2-FMN, and were confirmed with SDS–PAGE followed by Coomassie staining.

ITC experiments were performed by injecting the dark state mutant C450A of PA-Rac1 (0.14 mM) or the lit state mutant I539E (0.13 mM) into the CRIB domain of PAK1 (10 μ M) using a Microcal VP-ITC calorimeter at 25 °C. Each titration consisted of 29 injections of 10 μ l of mutants of PA-Rac1. The baseline of each titration was determined and subtracted from all of the data points. Titration data for the heat change per injection were fitted to a one-site binding model using Origin software (OriginLab).

Cell culture. HeLa, HEK293 (ATCC) and parental MEF/3T3 Tet-Off cells (Clontech) were maintained in DMEM containing 10% FBS following the supplier's culturing instructions. Stable MEF lines were passaged with addition of 1 ng ml^{−1} doxycycline, sufficient to suppress protein expression under the Tet-CMV promoter. Doxycycline was removed 24 h before live cell imaging. It was important to control expression level because small amounts of Rac activity from PA-Rac were apparently present before irradiation, as evidenced by increased ruffling at high expression levels. This was probably due to the equilibrium amount of active Rac present in the dark state. For constructs tagged with fluorescent proteins, expression level could be roughly approximated as proportional to brightness/unit area, enabling use of cells with similar expression.

Live cell microscopy. Cells for live cell imaging were seeded on coverslips coated with $5 \mu\text{g ml}^{-1}$ fibronectin in Ham's F-12K medium free of Phenol red and containing 2% fetal bovine serum (FBS). Coverslips were mounted in an Attofluor live cell chamber (Invitrogen) placed in a microscope stage with a heated stage adaptor (Warner). Initial characterization and photoactivation of PA-Rac1, diffusion studies by FRAP and PA-GFP, and protrusion/retraction analyses were carried out using an Olympus FluoView 1000 confocal scanning microscope system equipped with a $\times 60$ 1.42 NA oil objective and lasers at 405, 458, 488, 515, 568 and 633 nm. Fluorescence images were acquired using 0.1% power from a 30 mW multi-line Ar ion laser (Olympus, minimum power possible without introducing a neutral density filter) and scanned at $2 \mu\text{s}$ per pixel. The illumination used for photoactivation of PA-Rac1 was between 0.1% power for $10 \mu\text{s}$ and 1% power for 1 ms in a $10\text{-}\mu\text{m}$ spot, at 458 nm. A more precise measure of the light dose used for activation was obtained by measuring the power after the objective using a power meter (Thorlabs), as described in Supplementary Fig. 7.

Biosensor images were acquired using an Olympus IX81-ZDC microscope equipped with a CoolSNAP HQ2 14-bit camera (Photometrics) and ET-CFP/YFP filters (Chroma) as described previously^{18,19}. Band-pass and neutral density filters were switched using motorized filter wheels under computer control (Ludl). CFP, FRET and YFP images were acquired using a 100 W Hg arc lamp with a 3% ND filter for 500, 250 and 250 ms, respectively. FRET/CFP ratio images were calculated after shading correction, background subtraction, binary masking and image registration using MetaMorph and MatLab software as described previously^{18–20}.

Our imaging conditions, exposure times and filters resulted in a donor (ECFP) bleed-through factor (into the FRET channel) of 0.25 and an acceptor (Citrine) bleed-through factor of 0.07, using the previously described approach^{18–20}. The E-FRET calculation was used as a measure of apparent FRET efficiency of the RhoA biosensor, producing a G factor of 1.57. Photobleaching-corrected FRET efficiency (E_{corr}) images were processed based on acceptor photobleaching as previously described²¹. Imaging YFP acceptor fluorescence of the RhoA biosensor was carried out using band-pass filters 500/20 (excitation) and 545/30 (emission).

Simultaneous photoactivation and biosensor imaging was achieved using the FRAP-3D instrument (MAG Biosystems), an illumination system with galvanometer-driven laser positioning. A laser source at 473 nm was incorporated into the illumination pathway using beam-combining mirrors (Chroma, or a 94%/6% Magic Mirror from Olympus).

Immunocytochemistry. MEF cells expressing mVenus-PA-Rac were plated onto coverslips with etched grids (Bellco) that were coated with fibronectin as described above. The cells were locally irradiated at 473 nm through a $\times 20$ phase-contrast objective. Immediately after protrusions were induced, the cells were fixed in 3.7% formalin (Sigma), permeabilized in 0.2% Triton X-100, incubated with anti-phospho-PAK antibody (Cell Signaling), and finally incubated with Alexa Fluor 594-conjugated secondary antibody (Molecular Probes). The numbers on the etched-grid coverslips were used to locate the immunostained cells that had been previously irradiated.

Measurement of protrusion length. Protrusions were analysed by drawing up to 8 lines per cell, perpendicular to the cell border and at least 45 degree apart (a few lines were dropped because of cell junctions). More than 14 cells were analysed

for each construct. The line scans were analysed as kymographs, oriented as shown in Supplementary Fig. 4. The protrusion length was defined as the maximal distance of the membrane from the left hand border of the kymograph after irradiation minus the average distance of the cell border from the left hand edge of the kymograph before irradiation.

Fluorescence recovery after photobleaching (FRAP). MEF cells expressing mVenus-tagged PA-Rac1 were irradiated (515 nm, 10% power for $100 \mu\text{s}$) at a $10\text{-}\mu\text{m}$ spot using the Olympus FluoView1000 confocal scanning microscope and laser detailed above. The intensity of mVenus fluorescence was monitored (515 nm, 0.1% power for $2 \mu\text{s}$) before and after photobleaching at 2-s intervals. The intensity of fluorescence within the bleached spot was normalized against the integrated intensity of the entire cell. The data were plotted and fitted to a single-exponential decay to obtain $t_{1/2}$ (τ_D). A $t_{1/2}$ of 12.1 s and F_i (immobile fraction) of $18.6 \pm 0.6\%$ were obtained ($n = 26$). The diffusion coefficient ($D = 0.55 \mu\text{m}^2 \text{s}^{-1}$) was estimated using $D = \omega^2/4\tau_D$ (where ω is the radius of the circular bleached spot), assuming exclusively free lateral diffusion^{22,23}.

PA-GFP tracking. MEF cells expressing mPA-GFP-tagged PA-Rac1 were irradiated (405 nm, 6 mW diode laser, 10% power for $10 \mu\text{s}$) in a $10\text{-}\mu\text{m}$ spot to switch on PA-GFP. The fluorescence of PA-GFP was monitored using a 488-nm laser, acquiring an image every 2 s. The intensities of fluorescence within the irradiated spot and an adjacent spot of the same size were quantified and normalized against the entire cell. The decay of activated PA-GFP fluorescence was fitted to a single exponential decay, yielding a $t_{1/2}$ of 14.6 s.

Inhibitor studies. MEF cells expressing mVenus-tagged PA-Rac1 were incubated with $1 \mu\text{M}$ myosin II inhibitor blebbistatin, $1 \mu\text{M}$ MLCK inhibitor ML-7, or $10 \mu\text{M}$ ROCK inhibitor Y-27632 (Calbiochem). Cells underwent the cell shape changes previously described²⁴ and then reached a stable state within 30 min. After this cells were irradiated with the 458-nm laser at a $10\text{-}\mu\text{m}$ spot to induce PA-Rac activation.

Protrusion/retraction analysis. Fluorescence images of MEF cells expressing mVenus-tagged PA-Rac1 and its mutants were masked based on intensity thresholding to produce binary images. Regions of protrusion were isolated by subtracting the binary image at a given time point from that at time 0. Conversely the binary images of retraction were obtained. Areas not part of protrusions or retractions, those that overlapped the time 0 image, were obtained by subtracting the above two images from that at time 0. Each binary image was assigned a different colour: red, protrusion; blue, retraction; green, area overlapping with time 0; and white, background. These operations were carried out using MetaMorph software.

Polarity index calculation. To obtain the polarity index ($\cos\theta$) of the migrating MEF cells, x and y coordinates were obtained for the centroid before movement (x_0, y_0), the centroid after movement (x_1, y_1), and for the centre of the irradiation spot (x_2, y_2), using MetaMorph software. The cos and sin values of the angles were obtained using simple triangle calculations. First two angles were defined using an arbitrarily selected horizontal line θ_1 = angle between the arbitrary line and the line from (x_0, y_0) to (x_1, y_1), and θ_2 = angle between the arbitrary line and the line between (x_0, y_0) and (x_2, y_2). The cos used to characterize polarity (see main text and Fig. 2d) was obtained using the following formula: $\cos\theta = \cos(\theta_1 - \theta_2) = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2$.

Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment

Zachary T. Schafer^{1†}, Alexandra R. Grassian^{1*}, Loling Song^{1*}, Zhenyang Jiang¹, Zachary Gerhart-Hines^{2,3}, Hanna Y. Irie¹, Sizhen Gao¹, Pere Puigserver^{1,2} & Joan S. Brugge¹

Normal epithelial cells require matrix attachment for survival, and the ability of tumour cells to survive outside their natural extracellular matrix (ECM) niches is dependent on acquisition of anchorage independence¹. Although apoptosis is the most rapid mechanism for eliminating cells lacking appropriate ECM attachment², recent reports suggest that non-apoptotic death processes prevent survival when apoptosis is inhibited in matrix-deprived cells^{3,4}. Here we demonstrate that detachment of mammary epithelial cells from ECM causes an ATP deficiency owing to the loss of glucose transport. Overexpression of ERBB2 rescues the ATP deficiency by restoring glucose uptake through stabilization of EGFR and phosphatidylinositol-3-OH kinase (PI(3)K) activation, and this rescue is dependent on glucose-stimulated flux through the antioxidant-generating pentose phosphate pathway. Notably, we found that the ATP deficiency could be rescued by antioxidant treatment without rescue of glucose uptake. This rescue was found to be dependent on stimulation of fatty acid oxidation, which is inhibited by detachment-induced reactive oxygen species (ROS). The significance of these findings was supported by evidence of an increase in ROS in matrix-deprived cells in the luminal space of mammary acini, and the discovery that antioxidants facilitate the survival of these cells and enhance anchorage-independent colony formation. These results show both the importance of matrix attachment in regulating metabolic activity and an unanticipated mechanism for cell survival in altered matrix environments by antioxidant restoration of ATP generation.

Epithelial cells are dependent on interactions with specific ECM components for survival, proliferation and differentiation functions⁵. Loss of matrix attachment of cultured epithelial cells activates a caspase-mediated apoptotic program known as anoikis². In glandular cancers, such as breast cancer, tumour cells are displaced from their normal matrix niches in the early stages of tumorigenesis when they proliferate into the lumen of hollow glandular structures. Filling of the luminal space is one of the hallmarks of early tumorigenesis.

Studies of luminal filling in three-dimensional (3D) structures of MCF-10A mammary epithelial cells and the developing mammary gland have demonstrated that apoptosis is involved in the clearance of centrally localized cells that lack matrix attachment; however, inhibition of apoptosis is not sufficient for the survival of matrix-deprived cells in the luminal space^{3,4,6,7}. Notably, several oncogenes, including ERBB2, have been shown to rescue cells from anoikis and prevent clearance of luminal cells in 3D acinar structures⁸, suggesting that these oncogenes prevent luminal clearance programs in addition to anoikis.

Another distinct feature of both matrix-detached MCF-10A cells and centrally located, matrix-deprived cells in MCF-10A acini is the

induction of autophagy^{9,10}. Because autophagy is a catabolic process commonly upregulated under conditions of starvation¹¹, the association of this process with matrix deprivation suggests that ECM attachment regulates metabolic activity as well as apoptosis. Here we determine the basis for the metabolic defects in matrix-detached cells, demonstrate that oncogenes can rescue these defects by restoration of glucose uptake and enhancement of antioxidant capacity, and unexpectedly find that antioxidants alone can rescue matrix-detached cells by restoration of ATP generation through fatty acid oxidation (FAO). Furthermore, we demonstrate that antioxidants promote anchorage-independent survival in two *in vitro* models of tumorigenesis.

To investigate whether ECM regulates cellular metabolism, we examined cellular ATP levels in MCF-10A cells cultured on adherent or non-adherent plates. We detected a substantial reduction in ATP in both MCF-10A (Fig. 1a) and primary human mammary epithelial cells (HMEC; Supplementary Fig. 2) that had been detached from the ECM for 24 h. We confirmed these results in lysates normalized for total protein (Fig. 1b) and by measuring the ATP/ADP ratio (Fig. 1c). The reduction of ATP in MCF-10A cells occurred between 12 and 24 h after ECM detachment (Supplementary Fig. 7), was not affected by inhibition of apoptosis (Fig. 1a and Supplementary Figs 4 and 5) or autophagy (Supplementary Fig. 6), and was rescued by the addition of reconstituted basement membrane (Supplementary Fig. 3). The expression of ERBB2 in these cells (ERBB2-MCF-10A) prevented the reduction in ATP after matrix detachment (Fig. 1a, c), suggesting that ERBB2 circumvents the matrix requirement for ATP production.

Given the evidence that EGFR is downregulated in detached cells and that its overexpression can rescue anoikis¹², we investigated the effect of ERBB2 on EGFR expression in detached cells. Indeed, ERBB2 expression caused a marked stabilization of EGFR after ECM detachment (Fig. 1d) that was critical for the rescue of ATP (Fig. 1e). This stabilization correlated with the maintenance of ERK1 and ERK2 (also known as MAPK3 and MAPK1, respectively) activation and an enhancement of PI(3)K–AKT signalling (Fig. 1d); however, inhibition of PI(3)K (but not ERK) abrogated the ERBB2 rescue of ATP (Fig. 1f, g). Furthermore, expression of a constitutively active variant of PI(3)K (PIK3CA(E545K)) or AKT (Myr-AKT1) is sufficient to rescue detachment-induced ATP (Fig. 1h). These results indicate that ERBB2 rescues the metabolic defect in matrix-detached cells by preventing the downregulation of EGFR and thus maintaining the activation of the PI(3)K pathway.

Given the critical role of PI(3)K–AKT in stimulating glucose transport^{13,14}, we investigated whether glucose uptake is altered in matrix-detached cells. Indeed, we found a marked deficiency in glucose uptake in detached cells that was rescued (in a PI(3)K-dependent

¹Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ³Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [†]Present address: Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA.

*These authors contributed equally to this work.

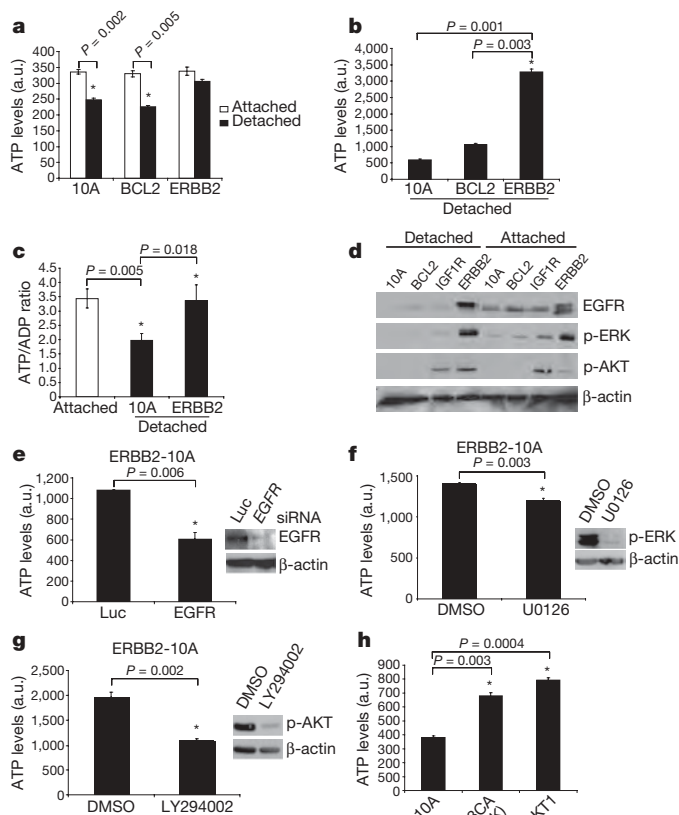


Figure 1 | Loss of matrix attachment causes reduction in cellular ATP that is rescued by ERBB2 through PI(3)K pathway activation. **a, b**, ATP was measured in the indicated cells 24 h after plating in adherent or non-adherent (poly-HEMA-coated) plates using the ATPlite assay (**a**) or the ATP determination kit (**b**). **c**, The ATP/ADP ratio was measured in the indicated cells. **d**, The indicated cells were immunoblotted for EGFR, phosphorylated (p)-ERK, p-AKT or β -actin. IGF1R cells were used as a positive control for p-AKT. **e**, ERBB2-MCF-10A cells were transfected with siRNAs targeting luciferase (Luc) or EGFR, and ATP levels were measured using the ATP determination kit. Knockdown was confirmed by immunoblot. **f, g**, Detached ERBB2-MCF-10A cells were treated with 10 μ M U0126 (ERK pathway inhibitor) (**f**) or 50 μ M LY294002 (PI(3)K inhibitor) (**g**) for 24 h, and ATP was measured using the ATP determination kit. Inhibition was confirmed by immunoblotting. **h**, ATP was measured in detached 10A cells or 10A cells expressing PIK3CA(E545K) or Myr-AKT1 using the ATP determination kit. All error bars represent s.d. ($n = 3$). Asterisks denote statistically significant change calculated using a two-tailed *t*-test. a.u., arbitrary units.

fashion, Supplementary Fig. 8) by ERBB2 expression (Fig. 2a). Furthermore, treatment of detached cells with methyl pyruvate, which provides substrates for the tricarboxylic acid (TCA) cycle, causes a substantial increase in ATP (Fig. 2b) that is blocked by the mitochondrial complex 1 inhibitor rotenone (Supplementary Fig. 9). This suggests that the mitochondria of detached cells retain the capacity to produce ATP. To determine whether the rescue of glucose uptake by ERBB2 is important for its ability to increase ATP levels in detached cells, we treated ERBB2-MCF-10A cells with 2-deoxyglucose (2DG), a glucose analogue that inhibits glycolysis. 2DG treatment completely abolished the rescue of ATP by ERBB2 (Fig. 2c), confirming the importance of glucose uptake in the rescue of ATP by ERBB2.

After cellular uptake, glucose can be further metabolised to generate ATP by glycolysis/oxidative phosphorylation, or it can be driven down the pentose phosphate pathway (PPP) by glucose-6-phosphate dehydrogenase (G6PD)¹⁵. Because the PPP is an important source of cellular NADPH (which provides reducing equivalents), we examined the effects of matrix detachment on ROS production. ECM detachment induced a significant increase in ROS (Fig. 2d) and decrease in

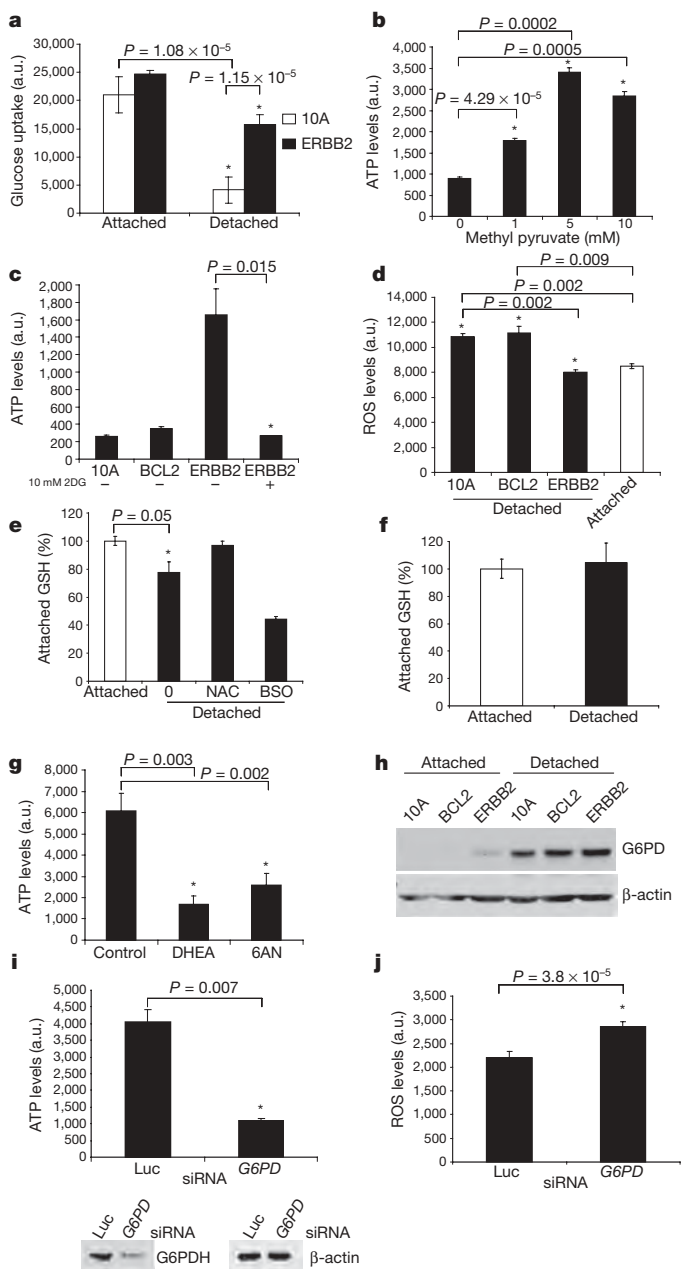


Figure 2 | Matrix detachment causes a reduction in glucose uptake and ERBB2 rescue of this defect is dependent on PPP flux. **a**, Glucose uptake was measured in the indicated cells using the Amplex Red assay. **b**, ATP was measured using the ATP determination kit after treatment (where indicated) of detached cells with methyl pyruvate. **c**, ATP was measured using the ATP determination kit after treatment (where indicated) of detached cells with 2DG. **d**, ROS was measured in attached and detached cells using DCF-DA. **e, f**, Reduced glutathione (GSH) was measured in 10A (**e**) or ERBB2-MCF-10A (**f**) cells. Results are shown as the percentage of the attached reduced GSH levels. NAC (1 mM) and BSO (1 mM; an inhibitor of glutathione synthesis) were used as a positive and negative control, respectively. **g**, ATP was measured using the ATP determination kit in detached ERBB2-MCF-10A cells after treatment (where indicated) with vehicle control, 150 μ M DHEA, or 150 μ M 6AN. **h**, The indicated cell lines were immunoblotted for G6PD after plating in either normal (attached) or poly-HEMA-coated (detached) plates. **i, j**, ATP (**i**) and ROS (**j**) were measured after siRNA-mediated knockdown of G6PD in detached ERBB2-MCF-10A cells. Knockdown was confirmed by immunoblotting. All error bars represent s.d. ($n = 3$). Asterisks denote statistical significance calculated using a two-tailed *t*-test.

reduced glutathione (Fig. 2e), both of which were reversed by ERBB2 (Fig. 2d, f). In support of the possibility that the lack of PPP flux is responsible for the increase in ROS, we found that the reduction in

glucose uptake precedes the increase in ROS levels (Supplementary Fig. 10).

Previous studies have shown that reducing ROS through the stimulation of PPP flux can promote cell survival^{16,17}; thus we proposed that PPP flux may be important for ERBB2 to rescue ATP levels in detached cells. Indeed, the treatment of detached ERBB2-MCF-10A cells with the PPP inhibitors dehydroisoandrosterone (DHEA) or 6-aminonicotinamide (6AN) abrogated the ability of ERBB2 to rescue ATP (Fig. 2g) and led to an increase in cellular ROS (Supplementary Fig. 11). Interestingly, we found that matrix-detachment caused a significant increase in G6PD protein (Fig. 2h), and short interfering RNA (siRNA)-mediated reduction of G6PD in detached ERBB2-MCF-10A cells caused a substantial decrease in ATP levels (Fig. 2i) and increase in ROS (Fig. 2j). The evidence that loss of matrix attachment causes an increase in ROS is consistent with studies in endothelial cells¹⁸ and in the developing mammary gland⁴.

The induction of ROS in suspended cells led us to investigate the effects of ROS neutralization on ATP in matrix-detached cells. Treatment of detached MCF-10A cells with the antioxidants *N*-acetyl-L-cysteine (NAC) or Trolox (a water-soluble vitamin E derivative) significantly increased ATP in detached cells (Fig. 3a), independent of any changes in glucose uptake (Fig. 3b). These data suggest that ROS inhibits ATP production through a metabolic pathway that could otherwise compensate for loss of glucose uptake.

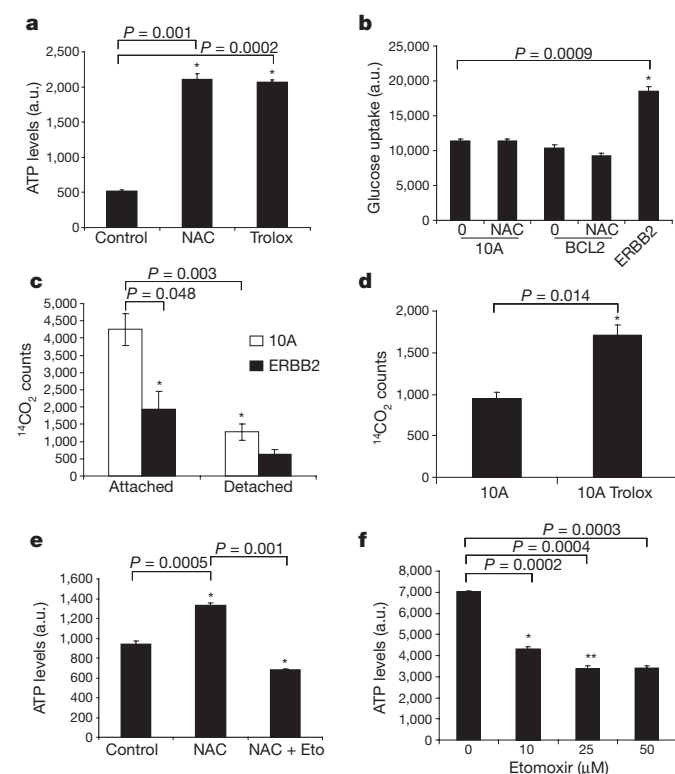


Figure 3 | Antioxidants rescue low ATP levels in detached cells by permitting fatty acid oxidation. **a**, MCF-10A cells were plated in poly-HEMA-coated plates and treated with control, 1 mM NAC or 50 μM Trolox, and ATP was measured using the ATP determination kit. **b**, Detached 10A or BCL2-MCF-10A cells were treated with either vehicle (0) or 1 mM NAC and glucose uptake was compared to ERBB2-MCF-10A cells using the Amplex Red assay. **c**, FAO was measured in parental or ERBB2-MCF-10A cells (either attached or detached). **d**, Detached MCF-10A cells were treated with either vehicle (10A) or 50 μM Trolox, and FAO was measured. **e**, Detached MCF-10A cells were treated with DMSO, 1 mM NAC, and/or 25 μM etomoxir (Eto) and ATP was measured using the ATP determination kit. **f**, Detached ERBB2-MCF-10A cells were treated with the indicated doses of etomoxir, and ATP was measured using the ATP determination kit. All error bars represent s.d. ($n = 3$). Asterisks denote statistical significance calculated using a two-tailed *t*-test.

It has previously been demonstrated that cancer cells deprived of glucose maintain ATP production through FAO¹⁹. Given that detached cells are glucose-deprived and antioxidants can promote ATP generation, we hypothesized that detachment-induced ROS could inhibit FAO. Indeed, we found that FAO was markedly reduced in detached MCF-10A cells (Fig. 3c) and Trolox treatment substantially increased FAO (Fig. 3d). In addition, treatment of antioxidant-supplemented cells with etomoxir, an inhibitor of fatty acid transport into the mitochondria¹⁹, prevented the rescue of ATP by antioxidants in detached cells (Fig. 3e). Furthermore, methyl malate treatment (which generates NADPH through an alternative pathway) increased ATP levels, lowered ROS levels, and caused an increase in FAO in detached MCF-10A cells (Supplementary Fig. 12a–c) suggesting that NADPH production is sufficient to increase FAO. Notably, ERBB2 reduced FAO in attached and suspended cells (Fig. 3c), consistent with previous reports showing PI(3)K–AKT-mediated reduction of FAO^{19,20}. However, treatment of detached ERBB2-MCF-10A cells with etomoxir caused a dose-dependent reduction in ATP levels (Fig. 3f), suggesting that the residual FAO significantly contributes to ATP production in these cells.

To extend our analysis to a model with more physiological relevance, we used the MCF-10A 3D cell culture model in which the inner, matrix-deprived cells show evidence of metabolic impairment (on the basis of upregulation of autophagy^{9,10}). Although it is not feasible to measure ATP directly in these structures, we used two-photon microscopy to examine the native fluorescence of NADH and NADPH (NAD(P)H) and obtain an assessment of metabolic differences between the inner and outer cells²¹. Because NADH is the principal electron donor in glycolytic and oxidative metabolism, the native fluorescence of NAD(P)H represents a non-invasive fluorescent reporter of the metabolic state²². Images of acinar structures on day 8 showed a dichotomy in NAD(P)H fluorescence intensity between the inner and outer cells of most structures, whereas most day 4 structures showed homogenous fluorescence (Fig. 4a and Supplementary Fig. 13). Furthermore, we found that ROS were detectable exclusively in the centrally localized cells of day 7 acini (Fig. 4c). Together, these results provide evidence that there are differences in metabolic activity and ROS accumulation in the matrix-deprived centrally localized cells of acini, as were observed in monolayer and suspension cultures (Figs 1–3). In addition, Trolox treatment prevented the dichotomy in NAD(P)H fluorescence in 3D culture (Fig. 4b and Supplementary Fig. 14), suggesting that ROS significantly contributes to the metabolic dichotomy between the inner and outer cells.

To understand the implications of these findings with regards to the clearance of cells from the luminal space, we studied the effects of antioxidant treatment on the survival of cells located in the centre of acini. We treated acini with either NAC or Trolox and monitored luminal filling over time. Notably, we found that NAC or Trolox treatment significantly reduced luminal clearance (Fig. 4d) independent of effects on caspase activation (Supplementary Fig. 15), suggesting that the elimination of ROS can preserve the viability of matrix-deprived cells in the centre of acini. By extrapolation to our studies showing ATP reduction and ROS generation in matrix-deprived cells, we speculate that the viability is a consequence of rescuing these metabolic impairments. Furthermore, the late clearing of acini observed in BCL2-MCF-10A cells is suppressed by Trolox (Fig. 4d). These results indicate that suppression of both metabolic impairments and apoptosis are required for anchorage-independent survival.

The evidence that antioxidants can rescue metabolic defects of ECM-deprived cells and promote luminal filling in MCF-10A acini raises the question of whether antioxidants could promote the transforming activity of mammary epithelial cells. To examine this, we assayed anchorage-independent colony formation in soft agar in MCF-10A cells expressing oncogenes that promote hyperproliferation (human papilloma virus E7 protein) and suppress apoptosis (BCL2). Neither E7 nor BCL2 expression rescued ATP in matrix detached cells

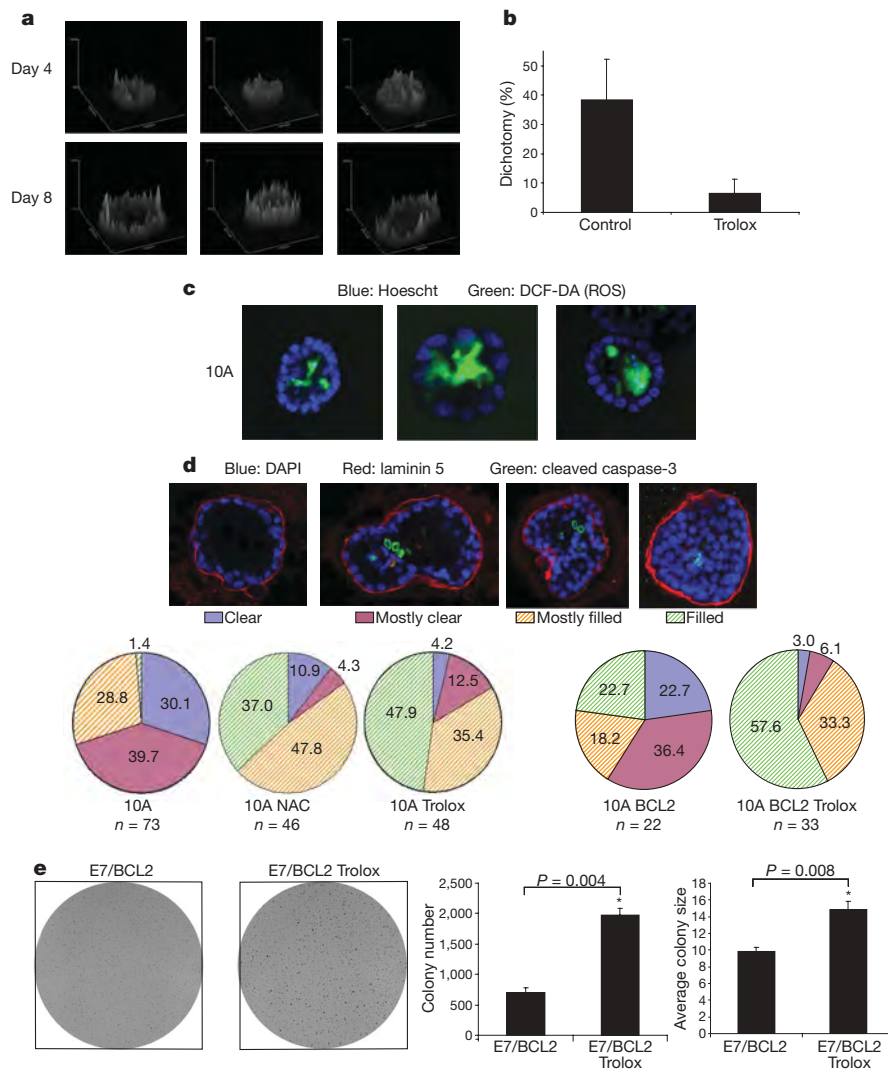


Figure 4 | Analysis of antioxidant effects on acinar morphogenesis and colony formation in soft agar. **a**, Native fluorescence of NAD(P)H was assayed by two-photon microscopy in MCF-10A acini. Surface intensity plots of three representative structures from day 4 or 8 are shown. **b**, Native fluorescence of NAD(P)H was assayed by two-photon microscopy in MCF-10A acini cultured in the presence or absence of 50 μ M Trolox (added daily). At day 8, structures were blindly scored for dichotomy in fluorescence intensity on the basis of evidence of a reduction in fluorescence intensity in the inner cells (see examples in Supplementary Fig. 14). Error bars represent s.d. of individual scorings ($n = 5$). **c**, ROS was measured in MCF-10A acini at day 7, by staining with 25 μ M DCF-DA (green) and Hoescht (blue). Three representative images are shown and 41% ($n = 73$) of all structures were positive for DCF-DA staining exclusively in the inner cells. NAC or Trolox

(data not shown) and the E7/BCL2 cells exhibited a relatively weak capacity to form colonies in this assay (Fig. 4e). Trolox treatment induced a substantial increase in the number and average size of colonies, suggesting that antioxidant treatment can enhance the transforming activity of cells that contain oncogenic insults. We observed similar results with ERBB2-MCF-10A cells (Supplementary Fig. 16) and with BT-474 cells—an ERBB2-expressing breast cancer cell line (Supplementary Fig. 17).

Our study highlights the possibility that glucose deprivation could lead to ROS production during tumorigenesis and force selection for alterations that allow escape from oxidative damage. Furthermore, we demonstrate that antioxidants rescue cells from the need for glycolysis through stimulation of FAO. It has been shown previously that stimulation of FAO through AMP-activated protein kinase (AMPK) or p53 (also known as TP53) is another mechanism for the survival of cells in the absence of glucose¹⁹. Thus, our study

treatment inhibited the DCF-DA staining in acini (data not shown). Original magnification, $\times 40$. **d**, Acini were formed using 10A or BCL2-MCF-10A cells (10A or 10A BCL2) and treated with 1 mM NAC or 50 μ M Trolox as indicated. At day 19 (10A) or day 33 (BCL2) acini were stained for laminin 5, cleaved caspase-3, and with the nuclear stain 4,6-diamidino-2-phenylindole (DAPI). Acini were then scored as described. Original magnification, $\times 40$. **e**, 10A cells expressing E7 and BCL2 were plated in soft agar and after 20 days, images were taken after iodonitroretazolium chloride (also known as INT-violet) staining. Colony number and average colony size were determined using ImageJ. Error bars represent s.d. ($n = 3$ unless otherwise indicated). Asterisks represent statistical significance calculated using a two-tailed *t*-test.

suggests that deprivation of matrix may limit glucose accessibility during tumorigenesis, and reveals strategies in which a tumour could escape from metabolic stress owing to these conditions by stimulating anchorage-independent glucose transport and/or eliminating ROS.

These data also demonstrate that antioxidants promote the survival of cells that lack attachment to the ECM and raise the possibility that antioxidants may have dichotomous activities with respect to tumorigenesis—that is, suppressing tumorigenesis by preventing oxidative damage to DNA^{23,24}, and promoting tumorigenesis by allowing survival of cells that are metabolically impaired (for example, in altered matrix environments). In support of this, expression of SOD2, a mitochondrial protein that reduces oxidative stress caused by respiratory chain leak, is increased in more advanced and higher grade mammary tumours^{25,26}. Furthermore, a recent study has revealed that enhanced PPP flux and increased antioxidant

capacity correlates with metastasis of breast cancer cells to the brain²⁷. Lastly, randomized trials have demonstrated both anti-neoplastic and neoplastic effects of antioxidants, with neoplastic effects associated with patients at higher risk owing to smoking and alcohol consumption²⁸, or patients undergoing chemo- or radiation therapy²⁹. Our work provides a biological rationale for these findings, as antioxidant activity may promote the survival of pre-initiated tumour cells in unnatural matrix environments and thus enhance malignancy.

METHODS SUMMARY

MCF-10A cells and their variants were all cultured as described at <http://brugge.med.harvard.edu/>. Other cell lines were cultured as described in the Supplementary Information. All assays on ECM-detached cells were carried out 24 h after plating on poly(2-hydroxyethyl methacrylate) (poly-HEMA)-coated plates unless otherwise noted. ATP assays were conducted using the ATPase assay (PerkinElmer), the ATP determination kit (Invitrogen), or the ATP/ADP Ratio Assay Kit (BioAssay Systems). Glucose uptake assays were performed using the Amplex Red Glucose Assay Kit (Invitrogen). ROS was measured using carboxy-H₂DCF-DA in detached/attached cells and in 3D culture. To measure reduced glutathione, we used chloromethylcoumarin (CMAC, Invitrogen). FAO was measured by monitoring the release of ¹⁴CO₂ after the addition of 1-¹⁴C-oleic acid. The 3D culture of mammary acini was completed according to the protocol at <http://brugge.med.harvard.edu/>. Native fluorescence of NADPH was measured using two-photon microscopy. Soft agar assays were performed in the presence or absence of antioxidants and colony formation/size was determined using ImageJ.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 March; accepted 6 July 2009.

Published online 19 August 2009.

- Simpson, C. D., Anyiwe, K. & Schimmer, A. D. Anoikis resistance and tumor metastasis. *Cancer Lett.* **272**, 177–185 (2008).
- Frisch, S. M. & Screaton, R. A. Anoikis mechanisms. *Curr. Opin. Cell Biol.* **13**, 555–562 (2001).
- Debnath, J. *et al.* The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell* **111**, 29–40 (2002).
- Mailleux, A. A. *et al.* BIM regulates apoptosis during mammary ductal morphogenesis, and its absence reveals alternative cell death mechanisms. *Dev. Cell* **12**, 221–234 (2007).
- Nelson, C. M. & Bissell, M. J. Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu. Rev. Cell Dev. Biol.* **22**, 287–309 (2006).
- Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
- Humphreys, R. C. *et al.* Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development* **122**, 4013–4022 (1996).
- Muthuswamy, S. K. *et al.* ErbB2, but not ErbB1, reinitiates proliferation and induces luminal repopulation in epithelial acini. *Nature Cell Biol.* **3**, 785–792 (2001).
- Fung, C. *et al.* Induction of autophagy during extracellular matrix detachment promotes cell survival. *Mol. Biol. Cell* **19**, 797–806 (2008).
- Mills, K. R. *et al.* Tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) is required for induction of autophagy during lumen formation *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 3438–3443 (2004).
- Mathew, R., Karantza-Wadsworth, V. & White, E. Role of autophagy in cancer. *Nature Rev. Cancer* **7**, 961–967 (2007).
- Reginato, M. J. *et al.* Integrins and EGFR coordinately regulate the pro-apoptotic protein Bim to prevent anoikis. *Nature Cell Biol.* **5**, 733–740 (2003).
- Czech, M. P. & Corvera, S. Signaling mechanisms that regulate glucose transport. *J. Biol. Chem.* **274**, 1865–1868 (1999).

- Elstrom, R. L. *et al.* Akt stimulates aerobic glycolysis in cancer cells. *Cancer Res.* **64**, 3892–3899 (2004).
- DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* **7**, 11–20 (2008).
- Bensaad, K. *et al.* TIGAR, a p53-inducible regulator of glycolysis and apoptosis. *Cell* **126**, 107–120 (2006).
- Boada, J. *et al.* Cells overexpressing fructose-2,6-bisphosphatase showed enhanced pentose phosphate pathway flux and resistance to oxidative stress. *FEBS Lett.* **480**, 261–264 (2000).
- Li, A. E. *et al.* A role for reactive oxygen species in endothelial cell anoikis. *Circ. Res.* **85**, 304–310 (1999).
- Buzzai, M. *et al.* The glucose dependence of Akt-transformed cells can be reversed by pharmacologic activation of fatty acid β -oxidation. *Oncogene* **24**, 4165–4173 (2005).
- DeBerardinis, R. J., Lum, J. J. & Thompson, C. B. Phosphatidylinositol 3-kinase-dependent modulation of carnitine palmitoyltransferase 1A expression regulates lipid metabolism during hematopoietic cell growth. *J. Biol. Chem.* **281**, 37372–37380 (2006).
- Bennett, B. D. *et al.* Quantitative subcellular imaging of glucose metabolism within intact pancreatic islets. *J. Biol. Chem.* **271**, 3647–3651 (1996).
- Chance, B., Cohen, P., Jobsis, F. & Schoener, B. Intracellular oxidation-reduction states *in vivo*. *Science* **137**, 499–508 (1962).
- Gao, P. *et al.* HIF-dependent antitumor effect of antioxidants *in vivo*. *Cancer Cell* **12**, 230–238 (2007).
- Narayanan, B. A. Chemopreventive agents alters global gene expression pattern: predicting their mode of action and targets. *Curr. Cancer Drug Targets* **6**, 711–727 (2006).
- Ivshina, A. V. *et al.* Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66**, 10292–10301 (2006).
- Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
- Chen, E. I. *et al.* Adaptation of energy metabolism in breast cancer brain metastases. *Cancer Res.* **67**, 1472–1486 (2007).
- Omenn, G. S. *et al.* Effects of a combination of β carotene and vitamin A on lung cancer and cardiovascular disease. *N. Engl. J. Med.* **334**, 1150–1155 (1996).
- Lawenda, B. D. *et al.* Should supplemental antioxidant administration be avoided during chemotherapy and radiation therapy? *J. Natl Cancer Inst.* **100**, 773–783 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Overholtzer, G. Mouneimne, M. Mazzone and C. Leung for critical reading of the manuscript and/or experimental assistance. We thank A. Mailleux, A. Kaanta, V. Schafer, A. Zhou, K. Simpson, and the members of the Brugge laboratory for experimental assistance, comments, and/or discussion. This work was supported by a grant from the National Cancer Institute (J.S.B.) and a grant from the National Institutes of Health (NIH) (P.P.). Z.T.S. is the recipient of a Ruth L. Kirschstein National Research Service Award (NRSA) for Postdoctoral Fellows from the National Cancer Institute; L.S. an NCI Mentored Quantitative Research Development Award (K25); A.R.G. a National Science Graduate Research Fellowship; and H.Y.I. an NCI K08 Award.

Author Contributions Z.T.S. and J.S.B. were responsible for the overall study design. Z.T.S., A.R.G., H.Y.I. and S.G. conducted experiments. L.S. and Z.J. conducted the experiments measuring native fluorescence of NADPH in 3D cell culture. Z.G.-H. and P.P. designed the fatty acid oxidation studies and Z.T.S. and Z.G.-H. conducted the fatty acid oxidation assays. Z.T.S. and J.S.B. drafted the manuscript and all other authors made revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.S.B. (joan_brugge@hms.harvard.edu).

METHODS

Cell culture. MCF-10A cells were cultured as described previously³⁰ and at <http://brugge.med.harvard.edu/>. MCF-10A cells expressing BCL2, ERBB2, E7/BCL2, PIK3CA(E545K), Myr-AKT1, IGF1R or GFP-LC3 were generated as described previously³. The pMSCV-Neo-based retroviral vectors encoding wild-type human *IGF1R* complementary DNA was a gift from R. Baserga. HMEC cells were immortalised with human TERT and cultured as described previously³¹. BT-474 cells (ATCC) were cultured in RPMI 1640 media plus 10% FBS and penicillin/streptomycin.

ATP assays. For the comparison of ATP levels in detached versus attached cells, the ATPase assay (PerkinElmer) was used. Cells were plated in 96-well poly-HEMA-coated (or uncoated) plates at a density of 13,333 cells per well. After 24 h, ATP assay was conducted according to the manufacturer's protocol. For the measure of ATP in detached cells (normalized by protein content), the ATP determination kit (Invitrogen) was used. Cells were plated in 6-well poly-HEMA-coated plates at a density of 400,000 cells per well. After 24 h, cells were lysed in 1% NP40 (plus protease inhibitors) and lysates were normalized by protein content using BCA Protein Assay (Pierce Biotechnology). Lysates were then tested for ATP levels according to manufacturer's protocol. The data from these assays show representative experiments from more than three independent replicates. For determination of the ATP/ADP ratio, the ADP/ATP Ratio Assay Kit (BioAssay Systems) was used according to manufacturer's instructions. The data from the ATP/ADP assay is presented as an average of the ATP/ADP ratio from several experiments.

3D cell culture. To generate acini, cells were grown in reconstituted basement membrane (Matrigel) as described previously³⁰ and according to the protocol at <http://brugge.med.harvard.edu/>. For antioxidant treatment, either NAC or Trolox was added to the Matrigel and to the overlay media. Immunofluorescence of acini was performed as described previously³⁰. The following primary antibodies were used for immunofluorescence: cleaved caspase-3 (Cell Signaling Technology) and laminin-5 (Millipore). DAPI (Sigma-Aldrich) was used to counterstain nuclei. For examination of luminal filling, acini were imaged using confocal microscopy to visualize the centre of each structure, and then were scored as clear (~90–100% clear), mostly clear (~50–90% clear), mostly filled (~10–50% clear), or clear (~0–10% clear). The figures including data from these assays show representative experiments from more than three independent replicates.

Reagents. The following reagents were used at the doses indicated and as described in the text/figure legends: methyl pyruvate (Sigma-Aldrich), 2-deoxy-D-glucose (Sigma-Aldrich), DHEA (EMD Biosciences), 6AN (Sigma-Aldrich), N-acetyl-L-cysteine (Sigma-Aldrich), Trolox (EMD Biosciences), etomoxir (Sigma-Aldrich), D-methyl malate (Sigma-Aldrich), Matrigel (BD Biosciences), poly(2-hydroxyethyl methacrylate) (poly-HEMA, Sigma-Aldrich), LY294002 (EMD Biosciences), U0126 (EMD Biosciences), staurosporine (Sigma-Aldrich), and DL-buthionine-(S,R)-sulphoximine (BSO, Sigma-Aldrich).

siRNA. The following siRNA SMARTpools (Dharmacon) were used: G6PD (M-008181-01), Beclin-1 (M-010552-00), ATG5 (M-004374-03), and EGFR (M-003114-01). The luciferase GL2 duplex (D-00110-01-20) was used as an siRNA control. Sequence information for each siRNA are as follows: G6PD SMARTpool, duplex 1, sense: GAGAGUGGGUUCAGUAUUU, antisense: 5'-PAUACUGGAAACCCACUCUCUU; duplex 2, sense: CAACAUCGCCUGC GUUAUCUU, antisense: 5'-PGAUAACGCAGCGCAUGUUGUU; duplex 3, sense: CGUGAGCCUGCGUAUUUUU, antisense: 5'-PAAUAACGCCAGG CCUCACGUU; duplex 4, sense: GACCUACGGCAACAGAUUUU, antisense: 5'-PUAUCUGUUGCCGUAGGUCAUU. Beclin-1 SMARTpool, duplex 1, sense: CUAAGGAGCUGCCGUUAUAUU, antisense: 5'-PUAUAACGGCAGC UCCUUAUU; duplex 2, sense: GGAUGACAGUAGAACAGUUAUU, antisense: 5'-PUAACUGUUCACUGUACCUU; duplex 3, sense: UAAGAUGGUCU GAAAUUUUU, antisense: 5'-PAAAUUUACAGCCAUUUUUU; duplex 4, sense: GCCAACAGCUUCACUCUGAUU, antisense: 5'-PUCAGAGUGAAGC UGUUGGCUU. ATG5 SMARTpool, duplex 1, sense: GGAAUAUCCUGC AGAAGAAUU, antisense: 5'-PUUCUUCUGCAGGAUAUCCUU; duplex 2, sense: CAUCUGAGCUACCCGGAUAUU, antisense: 5'-PUAUCCGGUGAGC UCAGAUUU; duplex 3, sense: GACAAGAAGACAUUAGUGAUU, antisense: 5'-PUCACUAAUGUCUUCUUGUCUU; duplex 4, sense: CAUUGGUUUU CUAUUUGAUU, antisense: 5'-PUCAAUAGCAAACCAUUGUU. EGFR SMARTpool, duplex 1, sense: GAAGGAAACUGAAUUCAAAUU, antisense: 5'-PUUUGAAUUCAGUUCCUUCUU; duplex 2, sense: GGAAUAUAGUAC UACGAAAUU, antisense: 5'-PUUUGUAGUACAUUAUUCUU; duplex 3, sense: CCACAAAGCAGUGAAUUAUU, antisense: 5'-PUAAAUUCACUGCU UUGUGGUU; duplex 4, sense: GUAACAAGCUCACGCAGUUUU, antisense: 5'-PAACUGCGUGAGCUUGUUAUU. Luciferase GL2 duplex, 5'-CGUACG CGGAAUACUUCGA dTdTdTdT GCAUGCGCCUUAUGAAGCU-5'.

For each transfection, 200 nM of siRNA was transfected into cells using oligofectamine (Invitrogen) according to manufacturer's protocol. Knockdown efficiency was examined after 48 h by western blotting as described later. For experiments involving siRNA in detached cells, cells were plated on poly-HEMA-coated plates 24 h after siRNA transfection and then assays were conducted 48 h after siRNA transfection. The figures including data using siRNAs show representative experiments from more than three independent replicates.

Western blotting. Cells were lysed in 1% NP40 on ice for 20 min. Lysates were spun at 16,000g at 4 °C for 30 min and normalized using a BCA Assay (Pierce Biotechnology). Lysates were then subjected to SDS-PAGE on polyacrylamide gels and transfer/blotting were performed as previously described⁴. The following antibodies were used for western blotting: G6PD (Abcam), Beclin-1 (Cell Signaling Technology), ATG5 (Cell Signaling Technology), cytochrome c (BD Biosciences), EGFR (Cell Signaling Technology), p-AKT (Cell Signaling Technology), p-ERK (Invitrogen) and β -actin (Sigma-Aldrich). The figures, including western blots, show representative blots from more than three independent experiments.

Cell death assays. Lysates preparation and the measurement of cytochrome c release were conducted as described previously³². For measuring viability by dye exclusion, detached or attached cells were stained with trypan blue and placed in a haemocytometer. Two-hundred cells were then counted and assessed as either trypan blue positive or negative.

Glucose uptake assay. For the analysis of glucose uptake, the Amplex Red Glucose Assay Kit (Invitrogen) was used. Cells were plated at a density of 13,333 cells per well in 96-well poly-HEMA-coated (or normal) plates. After 24 h, media was collected and diluted 1:4,000 in water. The amount of glucose in the media was then determined using the Amplex Red Assay according to the manufacturer's instructions. Glucose uptake was determined by subtracting the amount of glucose in each sample from the total amount of glucose in the media (without cells). The data from these assays show representative experiments from more than three independent replicates.

ROS and glutathione assays. Cells were plated at a density of 13,333 cells per well in 96-well poly-HEMA-coated (or normal) plates. After 24 h, carboxy-H₂DCFDA (Invitrogen) was added to each well at a concentration of 10 μ M and mixed well. Carboxy-H₂DCFDA is a cell-permeant indicator for reactive oxygen species that is retained in the cell after deacetylation and non-fluorescent until oxidation occurs within the cell. After 2–3 h, absorbance was monitored using a fluorimeter. The figures including data from these assays show representative experiments from more than three independent replicates. Alternatively, reduced glutathione (GSH) was measured as a proxy for ROS using CMAC (Invitrogen)³³. CMAC was added to detached or attached cells at 40 μ M and fluorescence was measured at an excitation of 360 nm and an emission of 460 nm. For the measurement of ROS in MCF-10A acini, the acini were stained with 25 μ M carboxy-H₂DCFDA and 1 μ M Hoescht 33342 for 1 h at 37 °C. Acini were then immediately imaged using confocal microscopy.

FAO assay. Cells were plated at a density of 100,000 cells per well in 12-well poly-HEMA-coated (or normal) plates. After ~3–4 h, 100 μ M oleic acid (Sigma-Aldrich) was added to each well. After 24 h, 1 μ l of 1 μ Ci μ l⁻¹ [¹⁴C]oleic acid (American Radiolabelled Chemicals) was added to each well and incubated at 37 °C for 1 h. To release ¹⁴CO₂, we then added 150 μ l of 3 M perchloric acid (Fisher Scientific) to each well and immediately covered each well with phenylethylamine (Sigma-Aldrich)-saturated Whatman paper. Plate was then incubated at room temperature overnight and subsequently Whatman paper was removed, placed into Ready-Safe Liquid Scintillation Fluid (Beckman Coulter), and ¹⁴C counts were read on scintillation counter. The figures including data from these assays show representative experiments from more than three independent replicates.

NAD(P)H fluorescence analyses. A two-photon excited fluorescence microscope was established on the basis of a Leica SP2 confocal microscope. It is equipped with a Ti:Sa laser (Spectral-Physics Broadband Mai Tai) with a tunable wavelength range of 710–990 nm, a pulse frequency of 80 MHz, and a pulse width (full-width at half-maximum (FWHM)) of 100 femtoseconds. The excitation wavelength was set to 730 nm. In the emission light path a SP700 short pass and BG39 filters were used to block any residual excitation photons. The fluorescence signal was filtered through a 505DCXR dichroic beam splitter (Chroma Technology) and wavelengths shorter than 505 nm were channelled to a Hamamatsu MCP-PMT detector. The effective spectral bandwidth of the detected fluorescence signal is 400–500 nm, corresponding to the maximum NADPH emission³⁴. A $\times 63$ (HCX PL APO 63X NA 1.2 W) water immersion objective lens was used for all the experiments. For each image, the emitted photons were collected for 5 min in the single photon counting mode (Becker & Hickl, SPC-830). All instrument settings and imaging parameters were maintained throughout the experiments. The images of photon counts are

displayed on the same intensity and spatial scales allowing for direct visual comparison.

MCF-10A cells were seeded on day 0 at a density of 5,000 cells per well in a coverslip-bottom 8-well chamber. On days 4 and 8, NAD(P)H fluorescence in 3D MCF-10A structures was imaged at the equatorial cross sections using two-photon fluorescence microscopy while being maintained on stage in a 37 °C chamber with 5% CO₂. For the antioxidant experiments (Fig. 4b) the two-photon excited native fluorescence images were acquired through non-descanned detection with a broadband emission filter (400–700 nm) and 730 nm excitation. Images were deconvolved by Huygens Professional (Scientific Volume Imaging). The images of more than 90 acinar structures were acquired on day 8 for either control or Trolox (50 µM) treatment. The randomized, blinded images were evaluated by five independent observers for the presence of NAD(P)H dichotomy (native fluorescence intensity difference between the inner and outer cells on the basis of images from the centre of each structure). For acini with filled lumen, an average from all five observers was derived from the positive scoring of a criterion. The standard deviation indicates the variation in a score among the five observers from this experiment.

Soft agar assays. Cells (4×10^4) were added to 1.5 ml (\pm antioxidants) of growth media plus 0.4% low-melt agarose (Sigma-Aldrich) and layered onto a 2-ml bed

of growth media plus 0.5% low-melt agarose. Cells were fed every 2–3 days with 1 ml of growth media (\pm antioxidants). At the indicated times, growth media was removed and viable colonies were stained with iodinitrotetrazolium chloride (Sigma-Aldrich). Colony number and colony size were determined using ImageJ. The figures including data from these assays show representative experiments from more than three independent replicates.

30. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**, 256–268 (2003).
31. Overholtzer, M. *et al.* Transforming properties of YAP, a candidate oncogene on the chromosome 11q22 amplicon. *Proc. Natl Acad. Sci. USA* **103**, 12405–12410 (2006).
32. Schafer, Z. T. *et al.* Enhanced sensitivity to cytochrome c-induced apoptosis mediated by PHAPI in breast cancer cells. *Cancer Res.* **66**, 2210–2218 (2006).
33. Sebastia, J. *et al.* Evaluation of fluorescent dyes for measuring intracellular glutathione content in primary cultures of human neurons and neuroblastoma SH-SY5Y. *Cytometry A* **51**, 16–25 (2003).
34. Vishwasrao, H. D., Heikal, A. A., Kasischke, K. A. & Webb, W. W. Conformational dependence of intracellular NADH on metabolic state revealed by associated fluorescence anisotropy. *J. Biol. Chem.* **280**, 25119–25126 (2005).

LETTERS

Direct activation of protein kinases by unanchored polyubiquitin chains

Zong-Ping Xia¹, Lijun Sun^{1,2}, Xiang Chen^{1,2}, Gabriel Pineda¹, Xiaomo Jiang¹, Anirban Adhikari¹, Wenwen Zeng¹ & Zhijian J. Chen^{1,2}

TRAF6 is a ubiquitin ligase that is essential for the activation of NF- κ B and MAP kinases in several signalling pathways, including those emanating from the interleukin 1 and Toll-like receptors^{1–3}. TRAF6 functions together with a ubiquitin-conjugating enzyme complex consisting of UBC13 (also known as UBE2N) and UEV1A (UBE2V1) to catalyse Lys 63-linked polyubiquitination, which activates the TAK1 (also known as MAP3K7) kinase complex^{4,5}. TAK1 in turn phosphorylates and activates I κ B kinase (IKK), leading to the activation of NF- κ B. Although several proteins are known to be polyubiquitinated in the IL1R and Toll-like receptor pathways, it is not clear whether ubiquitination of any of these proteins is important for TAK1 or IKK activation. By reconstituting TAK1 activation *in vitro* using purified proteins, here we show that free Lys 63 polyubiquitin chains, which are not conjugated to any target protein, directly activate TAK1 by binding to the ubiquitin receptor TAB2 (also known as MAP3K7IP2). This binding leads to autophosphorylation and activation of TAK1. Furthermore, we found that unanchored polyubiquitin chains synthesized by TRAF6 and UBC13 (also known as UBE2D3) activate the IKK complex. Disassembly of the polyubiquitin chains by deubiquitination enzymes prevented TAK1 and IKK activation. These results indicate that unanchored polyubiquitin chains directly activate TAK1 and IKK, suggesting a new mechanism of protein kinase regulation.

To investigate the biochemical mechanism of TAK1 regulation, we reconstituted TAK1 activation *in vitro* using purified proteins, including E1, ubiquitin, UBC13–UEV1A (E2), TRAF6 (E3) and the TAK1–TAB1 (also known as MAP3K7IP1)–TAB2 kinase complex (Fig. 1a, see Supplementary Information for the source and purification of the proteins). These proteins were incubated together with an ATP buffer, and the activation of TAK1 was analysed by immunoblotting with an antibody specific for TAK1 or TAK1 phosphorylated at Thr 187 (ref. 6). As shown in Fig. 1b, in the presence of E1, UBC13–UEV1A, TRAF6 and ubiquitin, a slower migrating form of TAK1, which could be detected with the phosphorylated-TAK1 antibody, was apparent (lane 4). Phosphorylation of TAK1 was not detectable in the absence of E1, UBC13–UEV1A or TRAF6, or in the presence of ubiquitin mutants containing a mutation at Lys 63, which prevented polyubiquitin (polyUb)-chain synthesis. The phosphorylated form of TAK1 was converted to unphosphorylated TAK1 by lambda protein phosphatase (Fig. 1c), confirming that the large mobility shift of TAK1 was indeed caused by phosphorylation. These results demonstrate that TRAF6-catalysed Lys 63 polyubiquitination is both necessary and sufficient to activate TAK1 *in vitro*.

Two reactions occurred in the reconstitution system: ubiquitination and kinase activation. To uncouple these two steps, we first carried out the ubiquitination reaction in the presence of E1, ubiquitin, UBC13–UEV1A and TRAF6, but in the absence of the TAK1 kinase complex

(Fig. 2a). The reaction mixtures were treated with *N*-ethylmaleimide (NEM)—an alkylating agent that inactivates E1 and E2. After NEM was quenched by dithiothreitol (DTT) and removed by buffer exchange, the ubiquitination mixtures were incubated with the TAK1 kinase complex in the presence of ATP. Remarkably, the ubiquitination product(s) synthesized in the first step was sufficient to activate the TAK1 kinase (Fig. 2a, lane 6). This activation was not caused by ubiquitination of the TAK1 kinase complex, because E1 and E2 had been inactivated by NEM in step 1 before they encountered the TAK1 complex in step 2. Inactivation of E1 and E2 by NEM was complete, because after E1 and UBC13–UEV1A were incubated with NEM, they could no longer function together with TRAF6 to activate TAK1 in the second step (lanes 7–10). The activation of TAK1 correlated perfectly with the formation of polyUb chains—the synthesis of

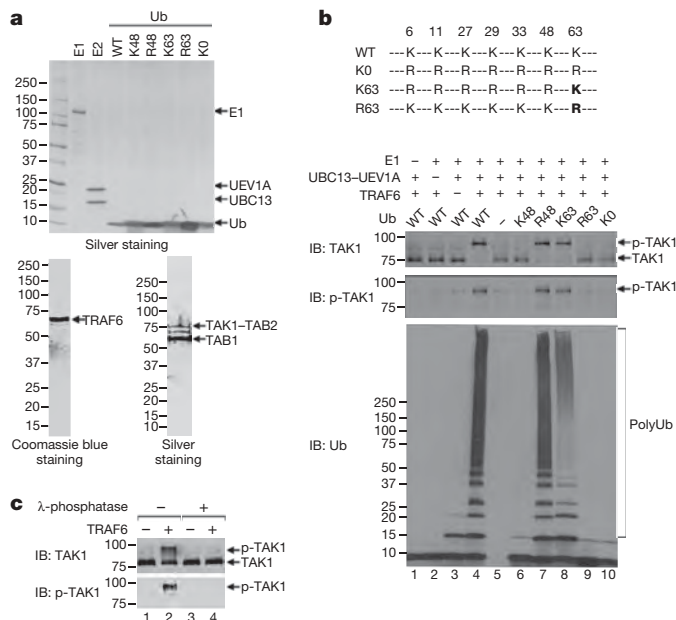


Figure 1 | *In vitro* reconstitution of TAK1 activation by TRAF6. **a**, Silver or Coomassie blue staining of purified proteins. Ub, ubiquitin. **b**, Top, diagram of ubiquitin lysine mutants used in the assays. K0 denotes no lysine residues. Bottom, the purified TAK1 kinase complex was incubated with E1, UBC13–UEV1A, TRAF6 and ubiquitin or its mutants in the presence of ATP for 1 h at 30 °C. Aliquots of the reaction products were analysed by immunoblotting (IB) with an antibody against TAK1, phosphorylated (p)-TAK1 or ubiquitin. **c**, TAK1 activated by TRAF6-catalysed ubiquitination as shown in **b** (lanes 3 and 4) was treated with lambda protein phosphatase and then analysed by immunoblotting. Numbers to the left of each gel denote kDa.

¹Department of Molecular Biology, ²Howard Hughes Medical Institute, University of Texas, Southwestern Medical Center, Dallas, Texas 75390-9148, USA.

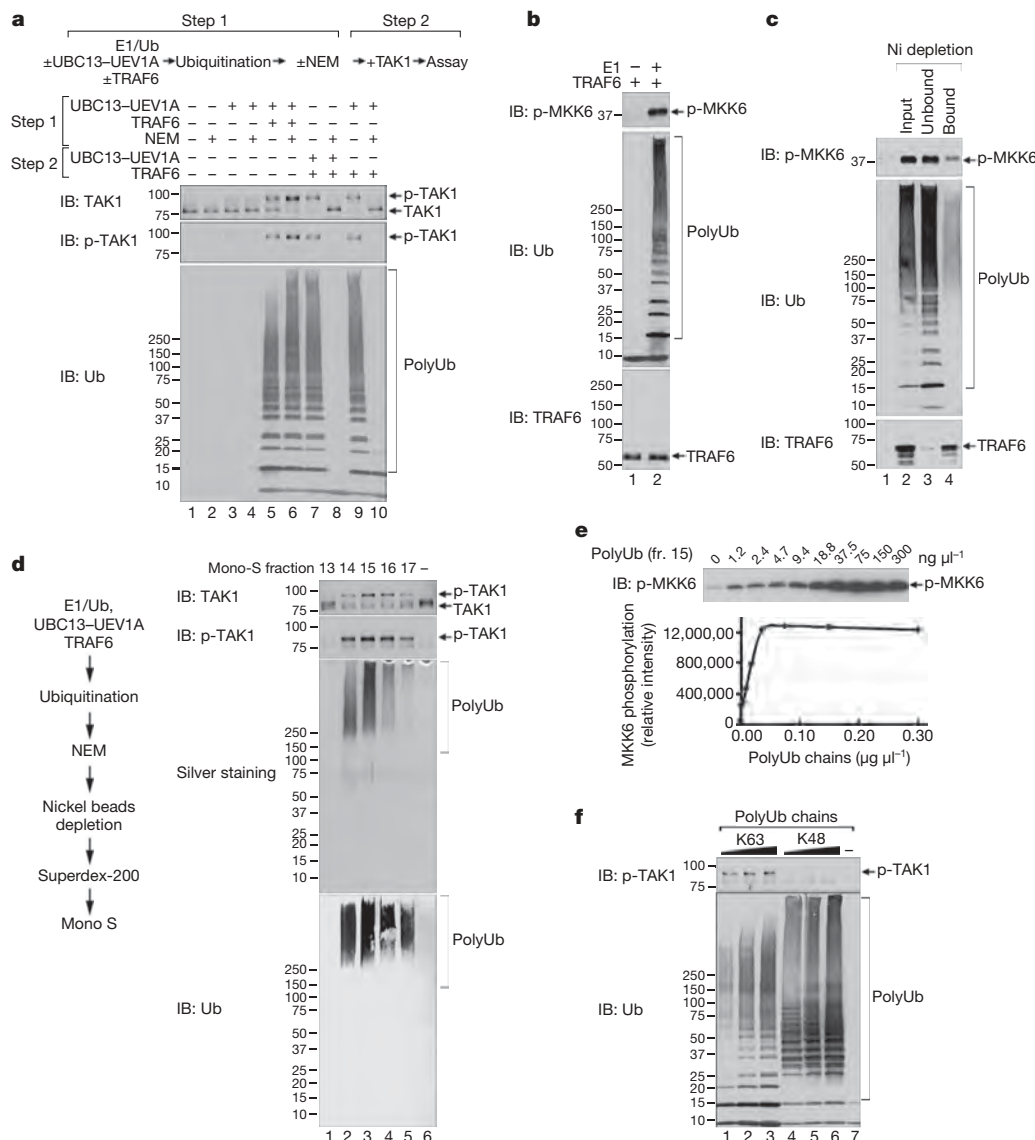


Figure 2 | Direct activation of TAK1 by unanchored polyubiquitin chains.

a, Ubiquitination reaction was carried out in the presence of E1, UBC13-UEV1A, TRAF6 and ubiquitin, then E1 and E2 were inactivated with NEM (step 1) before incubation with the TAK1 complex in the presence of ATP to measure TAK1 activation (step 2). **b**, Ubiquitination reactions were carried out as in step 1 in **a**, then incubated with TAK1 and its substrate MKK6(Lys82Ala). **c**, Ubiquitination mixtures were incubated with nickel-NTA

beads, then the bound and unbound fractions were tested for stimulation of TAK1. **d**, Left, the procedure for purification of TAK1 activator. Right, the Mono-S fractions were analysed for TAK1 activation and by silver staining and immunoblotting. **e**, Different concentrations of purified polyubiquitin chains (Mono-S fraction (fr. 15) in panel **d**) were tested for TAK1 activation. **f**, Lys 63- and Lys 48-linked polyubiquitin chains were tested for TAK1 activation. Numbers to the left of each gel denote kDa.

which required both UBC13-UEV1A and TRAF6. These results indicate that the polyubiquitination reaction in the first step produces an 'activator' that directly activates the TAK1 kinase complex.

A likely candidate that acts as the TAK1 activator is ubiquitinated TRAF6 (refs 4, 7). However, consistent with a previous report⁸, ubiquitinated TRAF6 was undetectable in the *in vitro* ubiquitination reactions containing varying concentrations of ubiquitin (Fig. 2b and Supplementary Fig. 2). Despite the absence of detectable TRAF6 ubiquitination, the ubiquitination reaction stimulated TAK1 to phosphorylate MKK6 (also known as MAP2K6), a physiological substrate (Fig. 2b, lane 2)⁴. Furthermore, when the ubiquitination reaction mixtures were incubated with nickel-NTA agarose to separate His-tagged proteins (including E1, UBC13-UEV1A and TRAF6) from the untagged protein (ubiquitin), most of the TAK1 stimulatory activity was present in the unbound fraction, which was largely depleted of TRAF6 (Fig. 2c, lane 3). Therefore, the TAK1 stimulatory activity correlated with the polyUb chains, but not with TRAF6.

Next, we tried to identify the TAK1 activator in an unbiased manner by biochemical fractionation of TRAF6-catalysed ubiquitination mixtures (see Methods and the diagram in Fig. 2d). The reaction mixtures were treated with NEM and then incubated with nickel-NTA agarose to deplete His-tagged proteins (E1, UBC13-UEV1A and TRAF6). The unbound material, which contained most of the TAK1 stimulatory activity, was further fractionated by gel filtration (Superdex-200) followed by cation exchange chromatography (Mono S). The fractions from the Mono-S column were assayed for their ability to activate TAK1 in the presence of ATP. The same fractions were also analysed by silver staining and immunoblotting with a ubiquitin antibody (Fig. 2d). Notably, the TAK1-stimulatory activity co-fractionated with high molecular mass polyubiquitinated species. Immunoblotting experiments with an antibody against TRAF6 or the His₆-tag showed no detectable signal (Supplementary Fig. 3), suggesting that the polyUb chains were not conjugated to any of these proteins. These 'unanchored' polyUb chains stimulated the TAK1

kinase to phosphorylate MKK6 in a dose-dependent and saturable manner (Fig. 2e). The concentration of polyUb chains required to achieve half-maximal activation of TAK1 was approximately $16 \text{ ng } \mu\text{l}^{-1}$. In contrast to Lys 63 polyUb chains, Lys 48-linked polyUb chains failed to activate TAK1 (Fig. 2f).

To demonstrate further that unanchored Lys 63 polyUb chains were responsible for TAK1 activation, we incubated the purified polyUb chains with the deubiquitination enzymes CYLD or isopeptidase T (also known as IsoT or USP5). CYLD is a tumour suppressor protein exhibiting deubiquitination enzyme (DUB) activity that specifically cleaves Lys 63-linked polyUb^{9,10}, whereas IsoT specifically

cleaves unanchored polyUb from the proximal end owing to its high-affinity binding to the C-terminal tail of ubiquitin¹¹. As shown in Fig. 3a, both CYLD and IsoT effectively converted Lys 63 polyUb chains into monoUb, and completely blocked the activation of TAK1. In contrast, a catalytically inactive CYLD mutant (Cys601Ser) was unable to cleave Lys 63 polyUb chains or inhibit TAK1. These results demonstrate that unanchored Lys 63 polyUb chains directly activate TAK1 *in vitro*.

To determine whether unanchored polyUb chains are synthesized in IL1 β -stimulated cells, we used a TAB2 antibody to immunoprecipitate the TAK1 complex and associated polyUb chains from HEK293

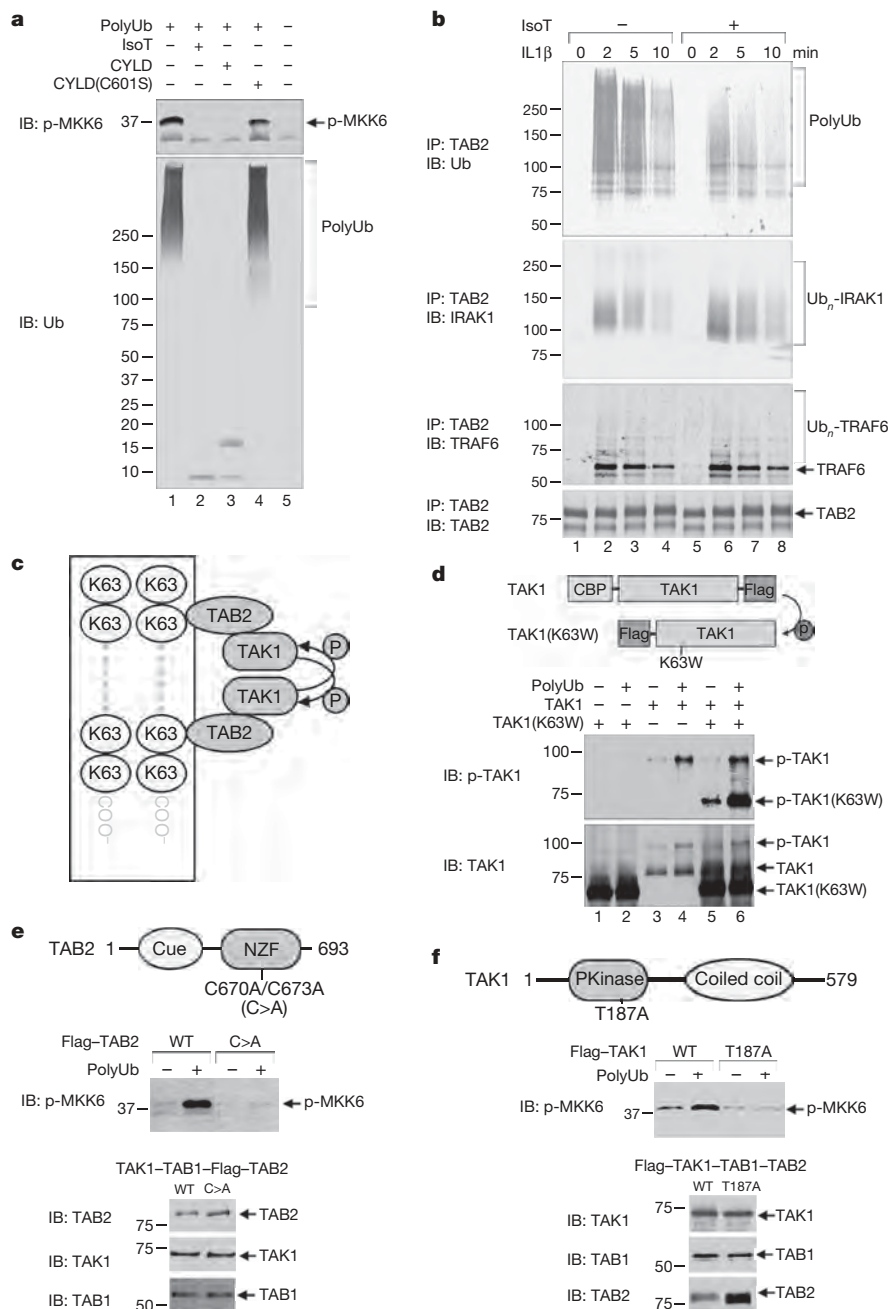


Figure 3 | IL1 β induces the synthesis of unanchored polyubiquitin chains to activate TAK1. **a**, Purified polyUb chains were pre-incubated with the indicated enzymes and then tested for TAK1 activation. **b**, HEK293T-IL1R cells were stimulated with IL1 β before cell lysates were immunoprecipitated (IP) with a TAB2 antibody. The precipitated materials were incubated with IsoT and then analysed by immunoblotting. Ub_n, polyubiquitinated. **c**, A model of polyUb-induced TAK1 autophosphorylation. **d**, Purified polyUb was incubated with the wild-type TAK1 and/or TAK1(Lys63Trp) complexes

in the presence of ATP. CBP, calmodulin-binding peptide. **e**, TAK1 complexes containing wild-type (WT) or a Cys670Ala/Cys673Ala (C>A) mutant of TAB2 were incubated with polyUb and MKK6(Lys82Ala) to measure TAK1 activation. Cue, coupling ubiquitin to endoplasmic reticulum degradation. **f**, Similar to **e**, except that the TAK1 complexes contained wild-type or Thr187Ala mutant of TAK1. Numbers to the left of each gel denote kDa.

cells stably expressing IL1R, which were stimulated with IL1 β for different lengths of time (Fig. 3b)¹². The TAB2-associated proteins were treated with IsoT, then immunoblotted with an antibody against ubiquitin, IRAK1 or TRAF6. Consistent with previous reports^{4,7,13,14}, IRAK1 and TRAF6 were rapidly polyubiquitinated after stimulation of cells with IL1 β , but the polyUb chains on these proteins were resistant to disassembly by IsoT, which only cleaves polyUb with an unanchored C-terminal tail. In contrast, immunoblotting with the ubiquitin antibody showed that most of the polyUb chains associated with TAB2 were cleaved by IsoT. Similarly, immunoprecipitation experiments with an antibody against NEMO (also known as IKBKG), which is known to bind to polyUb chains to mediate IKK activation^{15,16}, showed that the polyUb chains associated with NEMO in the IL1 β -stimulated cells were sensitive to IsoT treatment (Supplementary Fig. 4). These results indicate that IL1 β induced rapid formation of polyUb chains that associated with TAB2 and NEMO, but were not conjugated to any target protein. Further experiments suggest that ubiquitination of IRAK1 or TRAF6 is dispensable for IKK activation (see Supplementary Information and Supplementary Fig. 5), consistent with a recent report showing that a TRAF6 mutant lacking all lysine residues fully rescued IL1 β -induced TAK1 and NF- κ B activation in TRAF6-deficient cells¹⁷.

One possible mechanism by which polyUb chains activate the TAK1 kinase complex is through the binding of polyUb to the TAB2 subunit; this binding may facilitate autophosphorylation of TAK1 at Thr 187, which is known to be important for TAK1 activation

(Fig. 3c)⁶. To test this possibility, we expressed and purified kinase complexes containing wild type or a kinase-dead (Lys63Trp) mutant of TAK1. The two forms of TAK1 differ in size owing to a calmodulin-binding peptide appended to the amino terminus of the wild-type TAK1, allowing these proteins to be distinguished when probed with a TAK1 antibody. We incubated the wild-type and catalytically inactive TAK1 complex, both of which contained endogenous TAB2, in the presence of polyUb and ATP, and then measured the phosphorylation of TAK1 by immunoblotting. As shown in Fig. 3d, both wild-type and mutant TAK1 were phosphorylated at Thr 187 in the presence of polyUb when they were incubated together. Because TAK1(Lys63Trp) cannot phosphorylate itself, its phosphorylation at Thr 187 must have been carried out by the wild-type TAK1, suggesting that the TAK1 complexes are brought into close proximity by binding to polyUb. Alternatively, polyUb binding may allosterically activate the TAK1 complex.

The model proposed in Fig. 3c predicts that TAK1 activation by polyUb requires an intact ubiquitin-binding domain of TAB2 as well as Thr 187 of TAK1. Indeed, mutations of two conserved cysteine residues (C > A) within the novel zinc finger (NZF)-type ubiquitin-binding domain of TAB2 (ref. 12) severely impaired the activation of TAK1 by polyUb (Fig. 3e). Similarly, the TAK1(Thr187Ala) mutant failed to phosphorylate MKK6 in the presence of polyUb (Fig. 3f).

Whereas TRAF6 did not undergo auto-polyubiquitination in the presence of UBC13–UEV1A *in vitro*, it was strongly polyubiquitinated in the presence of UBCH5C (Fig. 4a). Immunoprecipitation of TRAF6

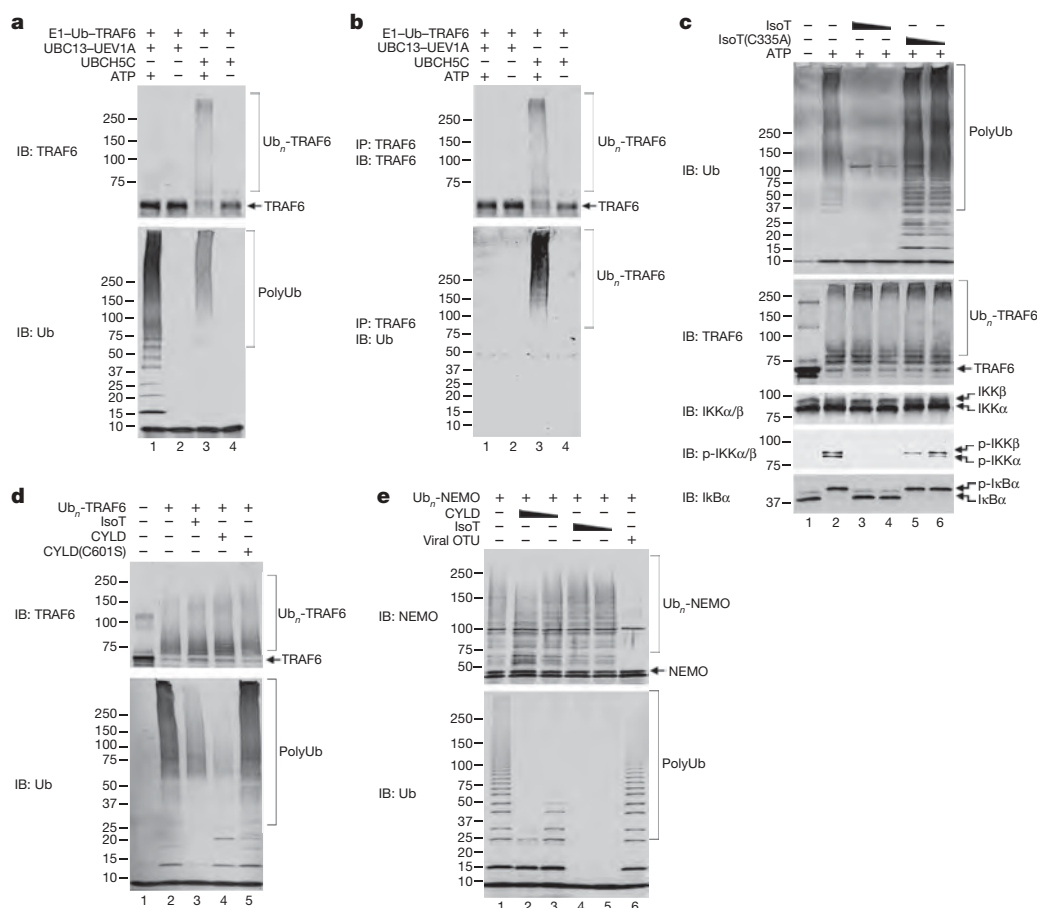


Figure 4 | Regulation of IKK by unanchored polyUb chains and CYLD.

a, Ubiquitination reactions containing TRAF6 and UBC13–UEV1A or UBCH5C were analysed by immunoblotting with the indicated antibodies. **b**, Aliquots of reaction mixtures from **a** were denatured by heating in 1% SDS, then diluted to 0.1% SDS before immunoprecipitation with a TRAF6 antibody. **c**, PolyUb synthesized in the presence of TRAF6 and UBCH5C was pre-incubated with IsoT or its mutant Cys335Ala, then incubated with the

IKK complex and Ikb α to measure IKK activation. **d**, Polyubiquitinated TRAF6 (Ub_n-TRAF6) was synthesized in the presence of UBCH5C, then incubated with the indicated DUBs before immunoblotting. **e**, Linear polyubiquitination of NEMO (Ub_n-NEMO) was carried out in the presence of UBCH5C and RNF31–RBCK1 (E3), then Ub_n-NEMO was incubated with the indicated DUBs followed by immunoblotting. Numbers to the left of each gel denote kDa.

under a denaturing condition confirmed that the polyUb chains were covalently attached to TRAF6 (Fig. 4b). Interestingly, the polyubiquitination products synthesized in the presence of UBCH5C preferentially activated IKK, whereas those synthesized by UBC13 activated TAK1 (Supplementary Fig. 6a, b). Treatments of polyUb chains with IsoT effectively cleaved the unanchored polyUb chains, but not polyubiquitinated TRAF6 (Fig. 4c). Notably, the IsoT treatment abrogated IKK activation, indicating that the UBCH5C-synthesized unanchored polyUb chains, but not ubiquitinated TRAF6, directly activated the kinase. A point mutation within a ubiquitin-binding domain of NEMO (Tyr308Ser) prevented IKK activation by unanchored polyUb chains *in vitro* (Supplementary Fig. 7), suggesting that the polyUb chains activate IKK by directly binding to NEMO. The direct activation of IKK by TRAF6 and UBCH5C may explain the observations that the ablation of UBC13 in some mouse cell types prevents the activation of TAK1 and MAP kinases, but not IKK¹⁸, and that TAK1-deficient cells retain residual IKK activity¹⁹.

Unlike UBC13–UEV1A, which catalyses specific Lys 63 polyUb chain synthesis^{5,20,21}, UBCH5C promotes the synthesis of polyUb chains of various linkages, including Lys 48 and Lys 63. Indeed, neither the mutation of Lys 48 nor Lys 63 to arginine prevented polyUb chain synthesis or IKK activation by TRAF6 and Ubc5, whereas ubiquitin mutants containing a single lysine at position 48 or 63 failed to support polyUb-chain synthesis or IKK activation (Supplementary Fig. 8). These results indicate that alternative ubiquitin linkages in the polyUb chains synthesized by TRAF6 and Ubc5 support IKK activation. A recent study suggests that linear ubiquitin chains, in which the N terminus of one ubiquitin is linked to the C terminus of the preceding ubiquitin, may mediate IKK activation²². However, we found no evidence that linear ubiquitin chains could efficiently activate TAK1 or IKK *in vitro* (see Supplementary Information and Supplementary Figs 9 and 10).

Surprisingly, like IsoT, CYLD specifically removed unanchored polyUb chains, but not polyUb chains conjugated to TRAF6 (Fig. 4d). To test whether CYLD could remove polyUb chains from another substrate, we carried out a ubiquitination reaction using UBCH5C and the RNF31 (also known as HOIP)–RBCK1 (also known as HOIL-1L) complex, which have been shown to function as E2 and E3, respectively, to catalyse linear polyubiquitination of NEMO²². Interestingly, CYLD and IsoT efficiently cleaved unanchored linear polyUb chains, but not the chains conjugated to NEMO (Fig. 4e). In contrast, a deubiquitination enzyme containing the ovarian tumour (OTU) domain of the Crimean-Congo hemorrhagic fever virus (CCHFV) large (L) protein removed polyUb chains from NEMO, but not unanchored linear polyUb chains²³ (Fig. 4e, lane 6). Further experiments indicated that the viral OTU, but not CYLD, was capable of cleaving Lys 63 polyUb chains from RIG-I (also known as DDX58), a viral RNA sensor²⁴ (see Supplementary Information and Supplementary Fig. 11). Taken together, these results indicate that CYLD preferentially cleaves unanchored polyUb chains *in vitro*. However, our data do not exclude the possibility that other proteins could assist CYLD in cleaving polyUb chains from protein targets or interfering with ubiquitination of these targets *in vivo*. Nevertheless, because the physiological function of CYLD in inhibiting TAK1 and IKK activation has been validated by abundant biochemical and genetic evidence⁹, our finding that CYLD preferentially cleaves unanchored polyUb chains further supports the role of these chains in protein kinase activation in the NF- κ B pathway.

In much the same way that second messengers such as cyclic AMP are generated to activate protein kinases after stimulation of cells, unanchored polyUb chains are rapidly synthesized in cells stimulated with IL1 β . Also like cAMP, which can be hydrolysed by phosphodiesterases, polyUb chains can be disassembled by specific deubiquitination enzymes, such as CYLD. The mechanism by which ubiquitin polymers activate protein kinases may be analogous to that used by polymeric nucleic acids, such as DNA and RNA, which bind to protein kinases containing nucleic-acid-binding domains or subunits (for

example, DNA-dependent protein kinase or protein kinase R). It is known that the binding of double-stranded RNA polymers to protein kinase R leads to its dimerization and autophosphorylation^{25,26}. Similarly, we propose that the binding of unanchored polyUb chains to TAB2 or TAB3 leads to the dimerization or oligomerization of the TAK1 kinase complex, and subsequent autophosphorylation of TAK1 at Thr 187, resulting in TAK1 activation. We also suggest that binding of unanchored polyUb chains to NEMO leads to autophosphorylation and activation of the IKK complex.

METHODS SUMMARY

Expression plasmids, proteins, antibodies and cell lines are described in the Methods. The TAK1 kinase complex was affinity purified from a HEK293T cell line stably expressing a TAP-tagged TAK1. The IKK complex was purified from HeLa S100 by affinity and conventional chromatography. Lys 63-linked polyUb chains were synthesized in a reaction mixture containing UBC13–UEV1A and TRAF6, and then purified by affinity and conventional chromatography. Lys 48-linked polyUb chains were synthesized using UBC3 (also known as CDC34) and the SKP1–CUL1–RBX1– β TrCP1 complex as the E2 and E3, respectively (β TrCP1 is also known as BTRC). The TAK1 and IKK activation assays were performed using polyUb chains as the activator. For deubiquitination experiments, polyUb chains were incubated with CYLD or IsoT. PolyUb chains induced by IL1 β were isolated by immunoprecipitation with an antibody against TAB2 or NEMO, incubated with or without IsoT, and then analysed by immunoblotting. A detailed description of the experimental procedures is provided in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 May; accepted 29 June 2009.

Published online 12 August 2009.

- Inoue, J., Gohda, J. & Akiyama, T. Characteristics and biological functions of TRAF6. *Adv. Exp. Med. Biol.* **597**, 72–79 (2007).
- Chen, Z. J. Ubiquitin signalling in the NF- κ B pathway. *Nature Cell Biol.* **7**, 758–765 (2005).
- Krappmann, D. & Scheidereit, C. A pervasive role of ubiquitin conjugation in activation and termination of I κ B kinase pathways. *EMBO Rep.* **6**, 321–326 (2005).
- Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
- Deng, L. *et al.* Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
- Singhirunnusorn, P., Suzuki, S., Kawasaki, N., Saiki, I. & Sakurai, H. Critical roles of threonine 187 phosphorylation in cellular stress-induced rapid and transient activation of transforming growth factor- β -activated kinase 1 (TAK1) in a signaling complex containing TAK1-binding protein TAB1 and TAB2. *J. Biol. Chem.* **280**, 7359–7368 (2005).
- Lamothe, B. *et al.* Site-specific Lys-63-linked tumor necrosis factor receptor-associated factor 6 auto-ubiquitination is a critical determinant of I κ B kinase activation. *J. Biol. Chem.* **282**, 4102–4112 (2007).
- Petroski, M. D. *et al.* Substrate modification with lysine 63-linked ubiquitin chains through the UBC13–UEV1A ubiquitin-conjugating enzyme. *J. Biol. Chem.* **282**, 29936–29945 (2007).
- Courtois, G. Tumor suppressor CYLD: negative regulation of NF- κ B signaling and more. *Cell. Mol. Life Sci.* **65**, 1123–1132 (2008).
- Komander, D. *et al.* The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Mol. Cell* **29**, 451–464 (2008).
- Reyes-Turcu, F. E. *et al.* The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* **124**, 1197–1208 (2006).
- Kanayama, A. *et al.* TAB2 and TAB3 activate the NF- κ B pathway through binding to polyubiquitin chains. *Mol. Cell* **15**, 535–548 (2004).
- Windheim, M., Stafford, M., Pegg, M. & Cohen, P. Interleukin-1 (IL-1) induces the Lys63-linked polyubiquitination of IL-1 receptor-associated kinase 1 to facilitate NEMO binding and the activation of I κ B α kinase. *Mol. Cell. Biol.* **28**, 1783–1791 (2008).
- Conze, D. B., Wu, C. J., Thomas, J. A., Landstrom, A. & Ashwell, J. D. Lys63-linked polyubiquitination of IRAK-1 is required for interleukin-1 receptor- and toll-like receptor-mediated NF- κ B activation. *Mol. Cell. Biol.* **28**, 3538–3547 (2008).
- Ea, C. K., Deng, L., Xia, Z. P., Pineda, G. & Chen, Z. J. Activation of IKK by TNF α requires site-specific ubiquitination of RI α and polyubiquitin binding by NEMO. *Mol. Cell* **22**, 245–257 (2006).
- Wu, C. J., Conze, D. B., Li, T., Srinivasula, S. M. & Ashwell, J. D. Sensing of Lys 63-linked polyubiquitination by NEMO is a key event in NF- κ B activation. *Nature Cell Biol.* **8**, 398–406 (2006).

17. Walsh, M. C., Kim, G. K., Maurizio, P. L., Molnar, E. E. & Choi, Y. TRAF6 autoubiquitination-independent activation of the NF κ B and MAPK pathways in response to IL-1 and RANKL. *PLoS One* **3**, e4064 (2008).
18. Yamamoto, M. *et al.* Key function for the Ubc13 E2 ubiquitin-conjugating enzyme in immune receptor signaling. *Nature Immunol.* **7**, 962–970 (2006).
19. Chen, Z. J., Bhoj, V. & Seth, R. B. Ubiquitin, TAK1 and IKK: is there a connection? *Cell Death Differ.* **13**, 687–692 (2006).
20. VanDermark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M. & Wolberger, C. Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* **105**, 711–720 (2001).
21. Hofmann, R. M. & Pickart, C. M. Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell* **96**, 645–653 (1999).
22. Tokunaga, F. *et al.* Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nature Cell Biol.* **11**, 123–132 (2009).
23. Frias-Staheli, N. *et al.* Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* **2**, 404–416 (2007).
24. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916–920 (2007).
25. Sadler, A. J. & Williams, B. R. Structure and function of the protein kinase R. *Curr. Top. Microbiol. Immunol.* **316**, 253–292 (2007).
26. Zhang, F. *et al.* Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop. *J. Biol. Chem.* **276**, 24946–24958 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C.-K. Ea for generating the HEK293 cell line stably expressing the TAP-tagged TAK1, J. Ashwell for the bacterial GST-Ub2 and GST-Ub3 expression plasmids, A. Garcia-Sastre for the expression vector encoding the viral OTU enzyme, CCHFV-L(1–169), and X. Li for the IRAK1-deficient HEK293 cells line. We also thank B. Skaug for critically reading the manuscript. This work was supported by grants from the National Institute of Health (RO1-AI09919 and RO1-GM63692) and the Robert Welch Foundation (I-1389). Z.J.C. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions Z.-P.X., L.S. and Z.J.C. designed the experiments, which were performed by Z.-P.X. and L.S., with assistance from X.C. G.P., X.J., A.A. and W.Z. contributed reagents. The manuscript was written by Z.J.C. and Z.-P.X.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.J.C. (Zhijian.Chen@UTSouthwestern.edu).

METHODS

Plasmids and cell lines. Expression plasmids for E1, ubiquitin, TRAF6, UBC13, UEV1A, TAK1, TAB2 and MKK6(Lys82Ala) have been described previously^{4,5}. N-terminal Flag-tagged IsoT, RNF31, RBCK1 and TRIM25 were constructed in pcDNA3.1. TAK1(Lys63Trp), TAK1(Thr187Ala), TRAF6-7KR(K > R substitutions at Lys91, 96, 104, 124, 133, 137 and 142), TAB2(Cys670Ala/Cys673Ala), and IsoT(Cys335Ala) were generated using QuikChange II XL kit (Stratagene). The bacterial expression plasmid for the viral OTU enzyme, CCHFV-L(1–169) was provided by A. Garcia-Sastre²³. All constructs were verified by DNA sequencing. IRAK1-deficient HEK293-IL1R and its parental cell line were provided by X. Li²⁷.

Protein expression and purification. Chromatography columns were from GE Healthcare unless otherwise indicated. E1 and TRAF6 were expressed in Sf9 cells as His₆-tagged proteins and purified using Ni-NTA agarose beads (QIAGEN). His₆-tagged UBC13, UEV1A, MKK6(Lys82Ala) and MBP-Ub-RIG-I(N) were expressed in *Escherichia coli* BL21/pLys and affinity purified. Glutathione S-transferase (GST)-tagged NEMO and the viral OTU enzyme, CCHFV-L(1–169) were similarly expressed in BL21/pLys and affinity purified followed by removal of GST with thrombin. Ubiquitin and its lysine mutants (no affinity tag) were expressed in a modified *E. coli* strain BL21(DE3)-pY2 (ref. 28), and purified through sequential chromatography on HiTrap-Q, HiTrap-SP and Superdex-75. Flag-tagged TRIM25, NEMO (wild-type and Tyr308Ser) were transiently expressed in HEK293 cells and purified using anti-Flag (M2) agarose beads (Sigma). GST-CYLD and its mutant (Cys601Ser) were expressed in Sf9 cells and purified by glutathione-sepharose chromatography.

Purification of the TAK1 complex. The TAK1 kinase complex was purified from a HEK293T cell line stably expressing TAK1, in which the N terminus of mouse TAK1 was fused to a tandem affinity tag consisting of protein A and calmodulin-binding peptide separated by a Tev protease cleavage site, and the C terminus was fused to a Flag tag. The cells were grown in DMEM media supplemented with 10% FBS. Cell extracts were prepared in buffer A (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 0.5% NP-40, 1 mM DTT, 1 mM PMSF and protease inhibitor cocktail (Roche)). The TAK1 complex, which contains endogenous TAB1 and TAB2, was purified using IgG-Sepharose and eluted by cleavage with Tev. The complex was further purified using anti-Flag (M2) agarose beads and eluted with the Flag peptide. The TAK1 complex containing wild-type TAB2 or its C > A mutant (Cys670Ala/Cys673Ala) was purified from HEK293T cells stably expressing the Flag-tagged TAB2 proteins using anti-Flag (M2) agarose beads. TAK1(Thr187Ala) was transiently expressed in HEK293 cells and then purified using its N-terminal Flag tag.

Purification of the IKK complex. HeLa S100 was loaded onto a HiTrap Q column and fractions containing the IKK activity were eluted with 200–300 mM NaCl. After precipitation with ammonium sulphate (15–30%) and dialysis, proteins were loaded onto a heparin column. Fractions eluted with 100–220 mM of NaCl were pooled and mixed with ATP-Sepharose beads²⁹. After washing with 1 M NaCl, bound materials were eluted with 100 mM ATP, and then loaded onto Superdex-200. Fractions containing IKK were used for assays.

For purification of endogenous IKK lacking NEMO, S100 from a NEMO-deficient murine pre-B cell line (1.3E2) was loaded onto a HiTrap Q column, and fractions containing IKK α/β were eluted with 200–300 mM NaCl. After precipitation with ammonium sulphate (15–30%) and dialysis, proteins were loaded onto Superdex-200. Fractions containing IKK α/β were used for assays.

Polyubiquitin chain synthesis and purification. Lys 63-linked polyUb chains were synthesized in a reaction mixture containing 0.2 μ M E1, 1 μ M UBC13–UEV1A, 0.4 μ M TRAF6 and 12.5 μ M ubiquitin in buffer B (50 mM Tris-HCl, pH 7.5, 2 mM ATP, 5 mM MgCl₂). After incubation at 30 °C for 1 h, the reaction was terminated by EDTA (10 mM). DTT (5 mM) was then incubated with the reaction mixture at room temperature for 15 min to release ubiquitin from E1 and E2, which were subsequently alkylated with 20 mM NEM (Sigma) for 15 min at room temperature. Finally, NEM was quenched with excess DTT (10 mM) and the reaction mixture was exchanged to storage buffer (20 mM HEPES-KOH, pH 7.4, 10% glycerol, 1 mM DTT, 0.4 mM PMSF) by repeated dilution and concentration using Amicon concentrators (Millipore). This mixture was incubated with nickel-agarose beads to deplete His₆-tagged proteins, and then polyUb chains were further purified by Superdex-200 followed by Mono-S chromatography using the SMART purification system.

PolyUb chains synthesized by UBCH5C and TRAF6 were prepared as above except that UBCH5C (E2, 0.1 μ M) and TRAF6 (0.4 μ M) were used. To test polyubiquitination of TRAF6, SDS was added to the reaction mixture to a final concentration of 1% and the sample was boiled for 4 min. After a tenfold dilution of SDS with buffer C (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% NP40), a TRAF6 antibody (0.2 μ g, Santa Cruz Biotech) and protein A/G agarose beads (5 μ l, Pierce) were added. The mixture was incubated for 2 h at room temperature and spun at 100g to collect protein A/G beads. The beads were extensively

washed with buffer C plus 0.1% SDS, and then proteins were eluted with 1% SDS and analysed by immunoblotting with a mouse antibody against TRAF6 or ubiquitin.

Lys 48-linked polyUb chains were synthesized as above except that Ubc3 (E2, 1 μ M) and the SKP1–CUL1–RBX1–BTRCP1 complex (E3, 0.4 μ M) were used to replace UBC13–UEV1A and TRAF6, respectively¹². Linear polyUb chains were synthesized in reactions containing UBCH5C (0.2 μ M), RNF31 (1 μ M) and RBCK1 (1 μ M). Lys 63-linked Ub3 and Ub4 were synthesized as previously described³⁰. Linear Ub2 and Ub3 were expressed as GST fusion proteins in *E. coli* (expression constructs provided by J. Ashwell¹⁶). After affinity purification, the GST tag was removed by digestion with thrombin, followed by absorption on a second glutathione column. Linear Ub2 and Ub3 were further purified by cation exchange chromatography using a HiTrapS column.

TAK1 and IKK activation assays. The TAK1 activation assay was carried out in 20 μ l of reaction mixture containing 0.1 μ M E1, 0.5 μ M UBC13–UEV1A, 0.2 μ M TRAF6, 12.5 μ M ubiquitin and the TAK1 complex (\sim 1 nM) in buffer B. The mixture was incubated for 1 h at 30 °C and analysed by immunoblotting with an antibody specific for phosphorylated Thr 187 of TAK1 (p-TAK1; Cell Signaling). In some assays, the reaction mixture contained the TAK1 substrate MKK6(Lys82Ala) (1.6 μ M) and the phosphorylation of MKK6 was analysed by immunoblotting using an antibody that recognizes phosphorylated Ser 207 of MKK6 (p-MKK6; Santa Cruz Biotech). To measure activation of TAK1 by purified polyUb chains, the TAK1 complex (\sim 1 nM) was incubated with polyUb (ranging from 1.2 to 300 μ g ml^{–1}) in a reaction mixture (20 μ l) containing 1.6 μ M MKK6(Lys82Ala) in buffer B. After incubation at 30 °C for 1 h, activation of TAK1 was analysed by immunoblotting with an antibody against p-TAK1 or p-MKK6.

To measure IKK activation by polyUb chains, the IKK complex (\sim 10 nM) was incubated with polyUb (\sim 150 μ g ml^{–1}) in a reaction mixture (10 μ l) containing \sim 50 nM Flag–IkB α in buffer B. After incubation at 30 °C for 1 h, IKK activation was analysed by immunoblotting with an antibody against p-IKK α/β (Cell Signaling) or IkB α (Santa Cruz Biotech).

Protein phosphatase and deubiquitination enzyme treatment. After TAK1 activation by TRAF6, ATP was removed using Zeba Desalt Spin column (Pierce), and the reaction mixture (20 μ l) was incubated with lambdaB protein phosphatase (10 U μ l^{–1}; New England Biolabs) at 37 °C for 30 min. Phosphorylation of TAK1 was then analysed by immunoblotting with a TAK1 antibody. For deubiquitination experiments, polyUb chains (3 μ g) were incubated with CYLD (200 nM) or IsoT (100 nM; Boston Biochem) in a reaction mixture (10 μ l) containing 20 mM Tris-HCl, pH 7.5, and 0.5 mM DTT for 1 h at 30 °C.

Ubiquitination and deubiquitination of TRAF6, NEMO and RIG-I. Polyubiquitination of TRAF6 was carried out in a reaction mixture containing 0.2 μ M E1, 0.1 μ M UBCH5C, 0.4 μ M TRAF6 and 12.5 μ M ubiquitin in buffer B at 30 °C for 1 h. Linear polyubiquitination of NEMO was carried out in a similar reaction except that NEMO (0.4 μ M) and RNF31–RBCK1 (0.4 μ M) were used in lieu of TRAF6. Lys 63 polyubiquitination of RIG-I was also carried out in a similar reaction except that 0.3 μ M of MBP-Ub-RIG-I(N), 0.4 μ M TRIM25 and 1 μ M UBC13–UEV1A were used to replace TRAF6 and UBCH5C. All reactions were terminated by EDTA (10 mM) before incubation with deubiquitination enzymes (IsoT, CYLD or viral OTU) at 37 °C for 1 h.

Isolation and detection of polyubiquitin chains induced by IL1 β . HEK293T-IL1R cells were stimulated with IL1 β for different lengths of time and then extracted with a lysis buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol, 0.5% NP-40, 20 mM 2-glycerolphosphate, 1 mM sodium orthovanadate, 0.5 mM DTT, and protease inhibitor cocktail. After centrifugation at 13,000g for 5 min, cell lysates (\sim 1 mg protein) were incubated with an antibody against TAB2 or NEMO on ice for 1 h, followed by a further 2-h incubation with protein A/G-Sepharose (Pierce) at 4 °C. Beads were washed extensively in lysis buffer and then resuspended in a buffer (10 μ l) containing 20 mM Tris, pH 7.5, 10 mM DTT, 0.1 mg ml^{–1} BSA and 10 μ g ml^{–1} of IsoT. After incubation at 30 °C for 1 h, the proteins on the beads were analysed by immunoblotting.

27. Li, X. *et al.* Mutant cells that do not respond to interleukin-1 (IL-1) reveal a novel role for IL-1 receptor-associated kinase. *Mol. Cell. Biol.* **19**, 4643–4652 (1999).
28. You, J., Cohen, R. E. & Pickart, C. M. Construct for high-level expression and low misincorporation of lysine for arginine during expression of pET-encoded eukaryotic proteins in *Escherichia coli*. *Biotechniques* **27**, 950–954 (1999).
29. Haystead, C. M., Gregory, P., Sturgill, T. W. & Haystead, T. A. γ -Phosphate-linked ATP-sepharose for the affinity purification of protein kinases. Rapid purification to homogeneity of skeletal muscle mitogen-activated protein kinase kinase. *Eur. J. Biochem.* **214**, 459–467 (1993).
30. Pickart, C. M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enzymol.* **399**, 21–36 (2005).

LETTERS

Structure of a tetrameric MscL in an expanded intermediate state

Zhenfeng Liu^{1,2}, Chris S. Gandhi¹ & Douglas C. Rees^{1,2}

The ability of cells to sense and respond to mechanical force underlies diverse processes such as touch and hearing in animals, gravitropism in plants, and bacterial osmoregulation^{1,2}. In bacteria, mechanosensation is mediated by the mechanosensitive channels of large (MscL), small (MscS), potassium-dependent (MscK) and mini (MscM) conductances. These channels act as 'emergency relief valves' protecting bacteria from lysis upon acute osmotic down-shock³. Among them, MscL has been intensively studied since the original identification and characterization 15 years ago⁴. MscL is reversibly and directly gated by changes in membrane tension. In the open state, MscL forms a non-selective 3 nS conductance channel which gates at tensions close to the lytic limit of the bacterial membrane. An earlier crystal structure at 3.5 Å resolution of a pentameric MscL from *Mycobacterium tuberculosis* represents a closed-state or non-conducting conformation^{5,6}. MscL has a complex gating behaviour; it exhibits several intermediates between the closed and open states, including one putative non-conductive expanded state and at least three sub-conducting states⁷. Although our understanding of the closed^{5,6} and open^{8–10} states of MscL has been increasing, little is known about the structures of the intermediate states despite their importance in elucidating the complete gating process of MscL. Here we present the crystal structure of a carboxy-terminal truncation mutant ($\Delta 95$ –120) of MscL from *Staphylococcus aureus* (SaMscL(CA26)) at 3.8 Å resolution. Notably, SaMscL(CA26) forms a tetrameric channel with both transmembrane helices tilted away from the membrane normal at angles close to that inferred for the open state⁹, probably corresponding to a non-conductive but partially expanded intermediate state.

The full-length SaMscL protein is 120-amino-acid-residues long and shares 40% and 51% homology with *Mycobacterium tuberculosis* MscL (MtMscL, also known as TbMscL) and the *Escherichia coli* MscL (EcMscL), respectively. Initial electrophysiological characterization of SaMscL (ref. 11) established that it exhibited characteristic membrane-tension-dependent MscL activity with significantly shorter open dwell times compared to EcMscL. Excised patches from azolectin liposomes containing SaMscL(CA26) show a stereotypical pressure-dependent increase in channel activity, similar to patches containing wild-type EcMscL, with a single channel conductance comparable to EcMscL (Fig. 1). The open state of SaMscL(CA26) appears more stable than that of wild-type SaMscL (Fig. 1c, d), suggesting a role for the C-terminal truncation in the gating behaviour change. Nevertheless, as shown in Supplementary Fig. 1, both full-length and SaMscL(CA26) are capable of rescuing a mechanosensitive channel knockout strain of *E. coli* cells from osmotic down-shock¹², indicating that they are both functional *in vivo*. Consequently, truncation of 26 residues from the C terminus of SaMscL does not impair its function, consistent with previous reports for EcMscL^{13,14}.

SaMscL(CA26) forms a tetramer in the crystal with the molecular symmetry axis coincident with the crystallographic four-fold axis (Supplementary Fig. 2), so that the asymmetric unit contains one subunit. The SaMscL(CA26) tetramer measures ~69 Å wide in diameter and ~37 Å thick (Fig. 2a, b), whereas the MtMscL pentamer, omitting C-terminal residues 102–151 which form a cytoplasmic helical bundle, measures ~52 Å wide and ~50 Å thick (Fig. 2c, d). Excluding the C-terminal residues that have been deleted in SaMscL(CA26), the major differences between the amino acid sequences of SaMscL and MtMscL are localized to the periplasmic ends of transmembrane helices TM1 and TM2 and their connecting loop. *In vitro* crosslinking experiments indicate that both CA26 and full-length SaMscL are tetrameric in detergent solution and that removal of the C-terminal domain does not change the oligomeric state of SaMscL (Supplementary Fig. 3). Although unexpected, it is not unprecedented that certain multimeric membrane proteins form distinct oligomers in different species. For example, the CorA Mg²⁺ transporter forms a pentamer in *Thermotoga maritima* and a tetramer

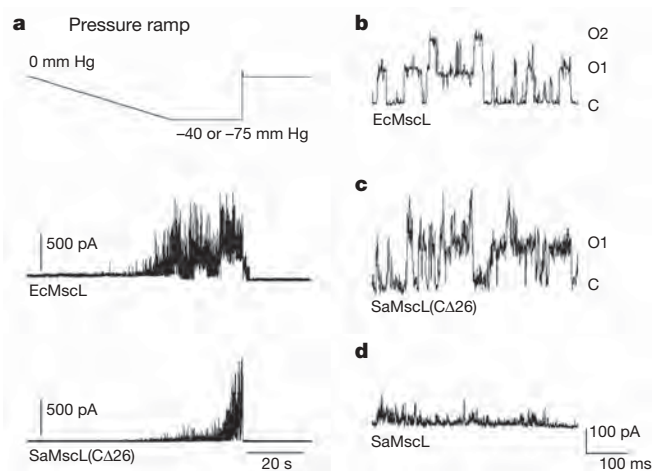


Figure 1 | SaMscL(CA26) is a pressure-sensitive channel. **a**, Pressure-sensitive response of EcMscL and SaMscL(CA26) proteins reconstituted into azolectin. Excised patches were held at +60 mV and pressure clamped from 0 to -40 mm Hg (EcMscL) or -75 mm Hg (SaMscL(CA26)). In both cases, channel activity steadily increases with greater negative pressure and decreases when the pressure is removed. **b–d**, Single channel recordings of EcMscL (**b**), SaMscL(CA26) (**c**) and wild-type SaMscL (**d**). EcMscL and SaMscL(CA26) channels open and close with discreet jumps in current. Calculated single channel conductances for EcMscL and SaMscL(CA26) are 2.9 and 2.6 nS, respectively. In contrast, single channel SaMscL openings flicker rapidly and are too brief to calculate a conductance, as observed previously¹¹. C denotes the closed state, whereas O1 and O2 indicate the opening of one and two channels, respectively.

¹Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. ²Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA.

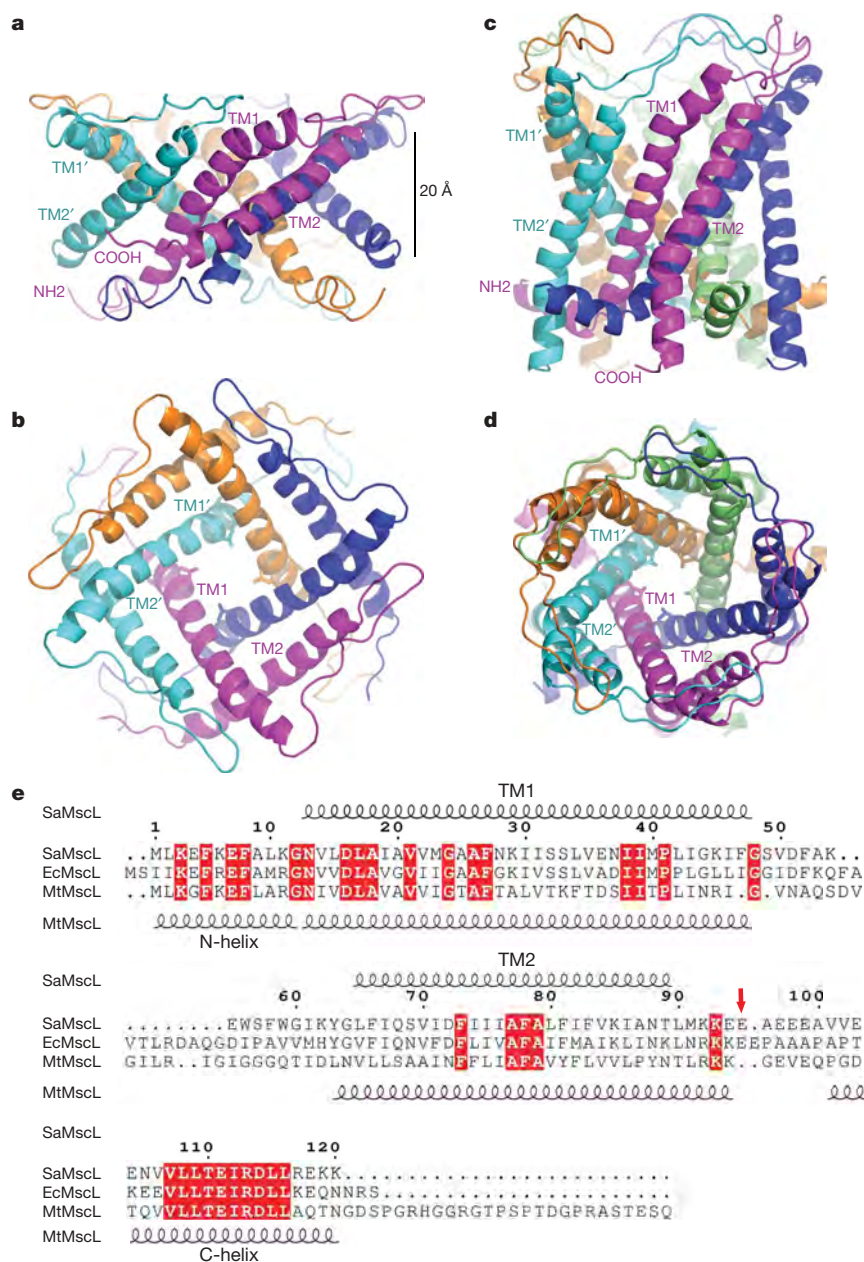


Figure 2 | Structures of SaMscL(CΔ26) and MtMscL. a, b, The crystal structure of tetrameric SaMscL(CΔ26) viewed along the membrane plane and the pore axis (from the periplasmic side), respectively. **c, d,** The membrane-spanning domain of the MtMscL structure (residues 1–101 of Protein Data

Bank set 2OAR) viewed as in **a** and **b**. Val 21 is presented as a stick model whereas the polypeptide chains are shown as ribbons. **e,** Structure-based sequence alignment of SaMscL, EcMscL and MtMscL. The red arrow indicates the truncation point (Glu 95) of the SaMscL(CΔ26) mutant.

in *Bacillus subtilis* and *E. coli*¹⁵, whereas the bacterial light-harvesting complex II may be either octameric or nonameric¹⁶. The ring of c subunits of the membrane-embedded F_0 -ATPase also exhibits variable stoichiometry¹⁷.

The secondary structure of the transmembrane region of SaMscL(CΔ26) resembles that of MtMscL (Fig. 2e) with a few exceptions. TM1 of SaMscL(CΔ26) begins at the cytoplasmic surface and extends from Asn 13 to Phe 47, kinking near Pro 41 on the periplasmic side. The polypeptide chain continues through the periplasmic loop from Gly 48 to Gly 64 and crosses the membrane again through TM2 from Leu 65 to Leu 90. TM1 has the same length as in MtMscL, whereas the periplasmic loop and TM2 are four residues and seven residues shorter, respectively, than the corresponding parts of MtMscL. The largest changes in the two structures are the first 12 residues (which adopt an irregular structure in SaMscL(CΔ26) but a short α -helix in MtMscL) and the detailed conformations of the periplasmic loop. In addition, the intrasubunit crossing angle between TM1 and TM2

changes from 134° in MtMscL to 111° in SaMscL(CΔ26) (Fig. 3a, b). An overlay of MtMscL and SaMscL(CΔ26) monomers on TM1 suggests an anticlockwise pivoting of TM2 away from the membrane normal and a translation towards the periplasmic surface (Fig. 3a). In both structures, TM1 and TM2 from the adjacent subunit (TM2') form an antiparallel pair of helices with crossing angle = 169° and 185° in SaMscL(CΔ26) and MtMscL, respectively. The TM1–TM2' pair of SaMscL(CΔ26) aligns well with that of MtMscL (root mean square deviation of α -carbon atoms = 1.5 Å) indicating that this pair might move as a rigid body during the conformational change (Fig. 3c and refs 18, 19). The extensive network of largely conserved hydrophobic residues at the TM1–TM2' interface presumably reflects the stability of this antiparallel helical pair (Supplementary Fig. 4).

The SaMscL(CΔ26) tetramer is ~13 Å thinner along the membrane normal than the membrane-spanning domain of MtMscL pentamer but up to 17 Å wider on the periplasmic surface. The periplasmic surface area of tetrameric SaMscL(CΔ26) is ~2,200 Å²,

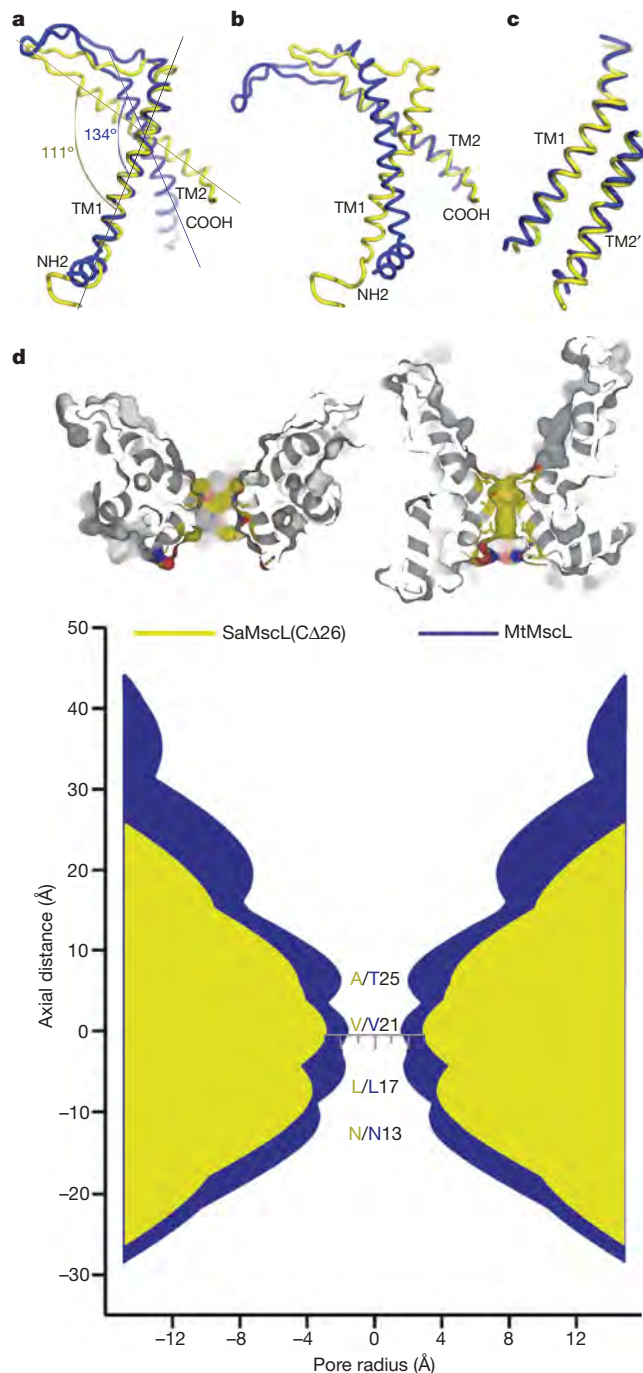


Figure 3 | Alignment of SaMscL(C Δ 26) (yellow) and MtMscL (blue). **a, b**, Intrasubunit conformational change visualized by aligning the two monomeric structures on TM1 and TM2, respectively. **c**, The superposition of the intersubunit TM1–TM2' pair from the two structures. **d**, The pore profiles of SaMscL(C Δ 26) and MtMscL along the central axes. The top two surface presentations are the sectional views of SaMscL(C Δ 26) (left) and MtMscL (right) with the atoms around the constriction coloured in yellow, red and blue for carbon, oxygen and nitrogen, respectively. In the bottom profiles, the regions corresponding to SaMscL(C Δ 26) and MtMscL are coloured yellow and blue, respectively. The residues lining the constriction are labelled in the same colour codes.

whereas that of pentameric MtMscL is $\sim 2,000 \text{ \AA}^2$. As such, the membrane-spanning region of SaMscL(C Δ 26) tetramer is expanded within the membrane plane and compressed along the pore axis as compared to the MtMscL pentamer (Fig. 2a–d). This expansion is also evident in the pore profiles of both channels (Fig. 3d). The constriction of SaMscL(C Δ 26) at Val 21 is widened to $\sim 6 \text{ \AA}$ diameter as compared to $\sim 3 \text{ \AA}$ in MtMscL. Ala 25 and Leu 17, positioned to

either side of Val 21 in the permeation pathway, move farther away from the central axis, opening the pore to $\sim 8 \text{ \AA}$ around those residues. The corresponding residues in MtMscL, Thr 25 and Leu 17, respectively, lie closer to the axis and hence the pore diameter in this structure narrows to $\sim 4 \text{ \AA}$ in this region. Although the constriction in SaMscL(C Δ 26) is wider than in MtMscL, theoretical studies indicate that hydrophobic pores with diameters less than 9 \AA and 13 \AA are impermeable to water and ions²⁰, respectively, implying that both structures are probably non-conducting because they each have narrow and hydrophobic pore constrictions. The expanded state of the channel of SaMscL(C Δ 26) reflects the higher tilt angles of TM1 and TM2, relative to the membrane normal. In SaMscL(C Δ 26), TM1 and TM2 are tilted 49° and 59° with respect to the pore axis, significantly larger than the $36\text{--}38^\circ$ for the transmembrane helices of MtMscL⁶, and close to the angles in an open-state model of EcMscL ($\sim 51^\circ$ for TM1/TM2)⁹. Consequently, the crossing angle between the axes of two adjacent TM1s in SaMscL(C Δ 26) (-63°) changes by $\sim -21^\circ$ relative to -42° in MtMscL. The combination of the constricted pore and expanded conformation suggests that SaMscL(C Δ 26) represents an intermediate state during the transition from the resting closed state to the sub-conducting or open states.

The SaMscL(C Δ 26) structure is clearly expanded in the crystal, but why was this conformation trapped? It is possible that removal of 26 residues from the carboxy terminus lowers the energy barrier for the transition from the closed to expanded state and crystal packing facilitates this specific expanded conformation. There are several lines of evidence consistent with this idea. Examination of the MtMscL structure suggests that the C-terminal bundle acts as a plug or pre-filter to the permeation pathway which may dissociate to permit the passage of large solutes, including small proteins^{21–23}, through the channel. This is consistent with recent electron microscopy work showing that a pore is present at the centre of a gain-of-function mutant (G22N) of EcMscL, whereas the cytoplasmic protrusion is absent²⁴. We suggest that the C-terminal helix bundle stabilizes the channel in the closed state and restrains the transmembrane helices from tilting into the expanded conformation. The dissociation of this bundle triggered by the mechanical force transmitted from the lipid bilayer removes the restraint, allowing the transmembrane helices to tilt and transit to the expanded state. Indeed, SaMscL(C Δ 26) reconstituted in liposomes is more stable in the open state than the wild-type channel, which displays fast flickering from the closed to open state¹¹ (Fig. 1c, d), suggesting that the truncation mutant may sample the expanded state more readily under crystallization conditions. The relatively small detergent micelle of *n*-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO)²⁵ might also favour the expanded state by exerting curvature stress on the channel.

To model the gating process of MscL, a series of structural models of tetrameric SaMscL from closed to open state was constructed using relationships between the helix tilting angles (η), interhelical-crossing angles (α) and the minimum pore radius (R) of oligomeric channels²⁶. We use equations $\cos \alpha = \cos^2 \eta + \sin^2 \eta \cos \theta$ and $R = d(\tan \eta \cot(\alpha/2) - 1)/2$, where $\theta = 360^\circ/N$ with $N = 4$ or 5 for a tetramer or pentamer and d is the helix diameter²⁶, as the basis for modelling a sequence of tetrameric structures with the α angles determined from various MscL structures in different states. Here we propose a two-step helix-pivoting model of SaMscL gating (see Supplementary movie). In this model, it is assumed that the TM1–TM2' pair rotates and shifts as a rigid body^{18,19}. The first pivoting step occurs during the transition from the closed state to a state near the expanded intermediate (Fig. 4a, b). TM1 pivots clockwise about Val 21 by $\sim 22^\circ$ and slides along the surface of TM2 (of the same subunit) from Gln 68 down to Ile 75. This induces TM2 to rotate anticlockwise about the pore axis and incline towards the membrane plane along with TM1. Increasing the TM1 tilt angle requires that the helices move farther apart to maintain a symmetric oligomer. As TM2 is in close interaction with TM1 from the next subunit, these movements are synchronized

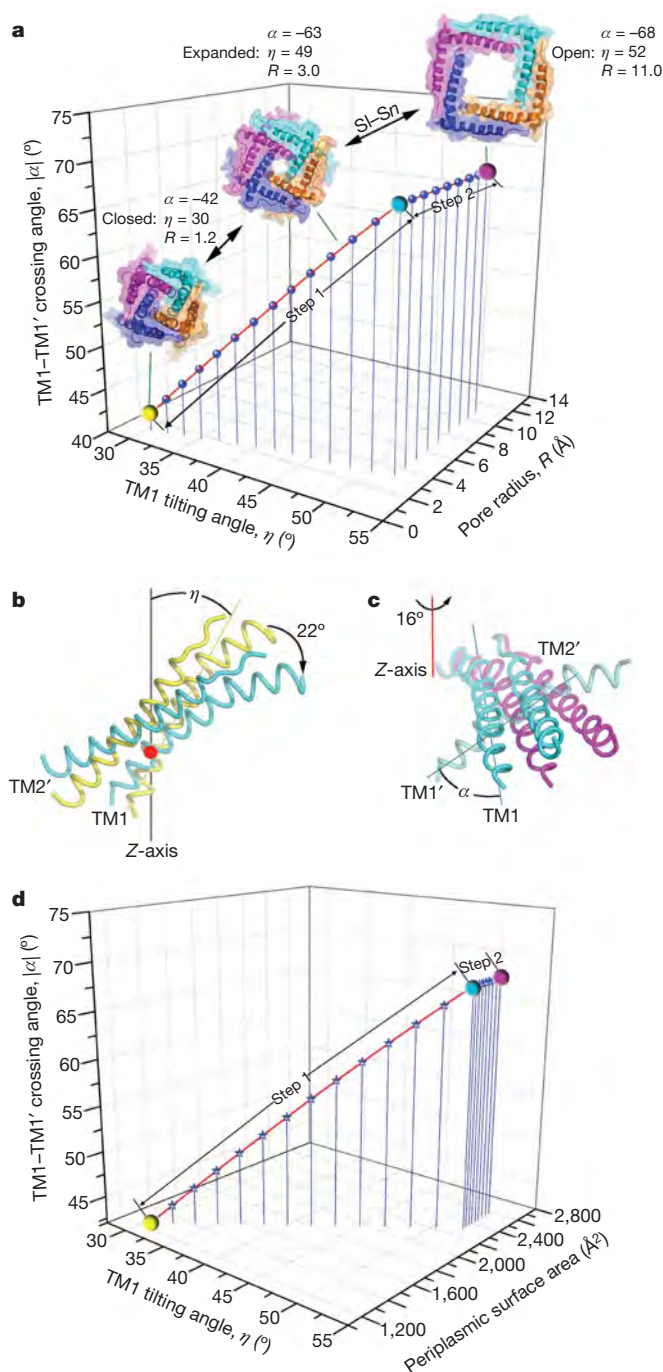


Figure 4 | The two-step helix-pivoting model of SaMscL gating. **a**, The relationship among TM1 tilting angle η , TM1–TM1' crossing angle α and the pore radius R at Val 21 derived from the gating models. The yellow, cyan and magenta spheres indicate the positions of the resting closed, turning-point intermediate and open states on the curve. S_1 – S_n denotes the multiple subconducting states between the expanded and open states. At step 1, η and α are greatly changed due to the helix tilting motion, whereas R is only slightly increased. As the helices swing away from the pore axis at step 2, R increases rapidly, whereas η and α remain constant. **b**, The first pivoting step of TM1–TM2' occurs about Val 21 (red point). The pair pivots as a rigid body rotating 22° from the resting closed (yellow, based on MtMscL structure^{5,6}) to the turning-point intermediate state (cyan, near the SaMscL(CA26) structure). **c**, The second pivoting step occurs about the red-line axis through Gly 48, from the intermediate to open state (magenta, based on the Sukharev and Guy model⁸). **d**, The relationship among η , α and the periplasmic surface area of the models. At step 1, the periplasmic surface area expands more drastically than at step 2. See also Supplementary movie.

throughout the tetramer. The crossing angle between two adjacent TM1s increases from -42° to -68° at this stage. This angle reaches -67° in a previous open-state model^{8,18} and -55° in an alternative model⁹, bracketing the value (-63°) observed in the expanded state structure, suggesting that it is probably near the edge of the open state. At the end of this first step, the periplasmic loop becomes fully stretched between intrasubunit TM1 and TM2, and if tilting continues, the channel would become thinner than the membrane.

The constraint of the membrane thickness leads to a proposed second pivoting step, defined as an anticlockwise rotation by $\sim 16^\circ$ of the TM1–TM2' pair about an axis running normal to the membrane and through Gly 48 near the C-terminal end of TM1 (Fig. 4a, c). The highly conserved Gly 48 (Fig. 2e) seems well suited to serve as the pivot for the outward swinging of the TM1–TM2' pair. This second-step action expands the pore diameter to 22 Å, compatible with the observed large conductance. In MscS, the pore-lining helix also pivots around a glycine residue and moves outwards from the central axis to expand the pore constriction²⁷. The periplasmic loop of MscL provides the flexibility for the substantial relative movements of the TM1 and TM2 observed within the same subunit, but also defines the steric limit of channel expansion in the fully open state. Previous functional studies on the periplasmic loop suggested that it may act as a spring, influencing channel kinetics and mechanosensitivity^{13,28–30}. During the first step of this transition, the periplasmic surface of SaMscL expands significantly while the channel constriction remains essentially closed until the second step takes place (Fig. 4a, d).

We have determined the crystal structure of a C-terminal truncation mutant of *S. aureus* MscL and demonstrated that SaMscL forms a tetrameric pressure-gated channel. This structure represents a likely expanded intermediate conformation of MscL between the resting closed state and the open state. A two-step helix-pivoting model of the gating mechanism has been proposed.

METHODS SUMMARY

The *Staphylococcus aureus* *mscL* gene was inserted into a pET15b vector with a stop codon at Glu 95 introduced by site-directed mutagenesis to generate the SaMscL(CA26) mutant. The channel was overexpressed in BL21-Gold (DE3) cells and purified in the detergent LDAO by immobilized metal affinity and gel filtration chromatographies. Crystals were grown using the hanging-drop vapour diffusion method. The structure was solved by the method of multiple isomorphous replacement with anomalous scattering.

Crosslinking was performed using *N,N*-dicyclohexylcarbodiimide at room temperature for 30 min to 1 h and the products were analysed by SDS-polyacrylamide gel electrophoresis.

The MJF465 (*mscL*[−], *mscS*[−], *mscK*[−]) (DE3) strain was used for the osmotic down-shock assay. Cells hosting EcMscL and an empty vector were included as positive and negative control sets, respectively. For the osmotic shock, the culture was diluted 1:500 into sterile double-distilled H₂O. For the mock shock, LB media with 500 mM NaCl was used instead of H₂O.

MscL protein was reconstituted into giant azolectin liposomes via the dehydration/rehydration method and assayed by patch-clamp electrophysiology. Patches containing at least 1–10 channels were excised and currents measured at various pressures. Recording and pipette solutions both contained 200 mM KCl, 40 mM MgCl₂, 5 mM HEPES, pH 7.2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 May; accepted 2 July 2009.

Published online 23 August 2009.

1. Kung, C. A possible unifying principle for mechanosensation. *Nature* **436**, 647–654 (2005).
2. Sukharev, S. & Corey, D. P. Mechanosensitive channels: multiplicity of families and gating paradigms. *Sci. STKE* **2004**, 1–24 (2004).
3. Booth, I. R., Edwards, M. D., Black, S., Schumann, U. & Miller, S. Mechanosensitive channels in bacteria: signs of closure? *Nature Rev. Microbiol.* **5**, 431–440 (2007).
4. Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R. & Kung, C. A large-conductance mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature* **368**, 265–268 (1994).
5. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* **282**, 2220–2226 (1998).

6. Steinbacher, S., Bass, R., Strop, P. & Rees, D. C. in *Current Topics in Membranes. Mechanosensitive Ion Channels, Part A* (ed. Hamill, O. P.) 1–24 (Academic, 2007).
7. Sukharev, S. I., Sigurdson, W. J., Kung, C. & Sachs, F. Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. *J. Gen. Physiol.* **113**, 525–540 (1999).
8. Sukharev, S., Betanzos, M., Chiang, C.-S. & Guy, H. R. The gating mechanism of the large mechanosensitive channel MscL. *Nature* **409**, 720–724 (2001).
9. Perozo, E., Cortes, D. M., Sompornpisut, P., Kloda, A. & Martinac, B. Open channel structure of MscL and the gating mechanism of mechanosensitive channels. *Nature* **418**, 942–948 (2002).
10. Anishkin, A. & Sukharev, S. State-stabilizing interactions in the bacterial mechanosensitive channel gating and adaptation. *J. Biol. Chem.* **284**, 19153–19157 (2009).
11. Moe, P. C., Blount, P. & Kung, C. Functional and structural conservation in the mechanosensitive channel MscL implicates elements crucial for mechanosensation. *Mol. Microbiol.* **28**, 583–592 (1998).
12. Levina, N. *et al.* Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J.* **18**, 1730–1737 (1999).
13. Blount, P., Sukharev, S. I., Schroeder, M. J., Nagle, S. K. & Kung, C. Single residue substitutions that change the gating properties of a mechanosensitive channel in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **93**, 11652–11657 (1996).
14. Häse, C. C., Le Dain, A. C. & Martinac, B. Molecular dissection of the large mechanosensitive ion channel (MscL) of *E. coli*: mutants with altered channel gating and pressure sensitivity. *J. Membr. Biol.* **157**, 17–25 (1997).
15. Niegowski, D. & Eshaghi, S. The CorA family: structure and function revisited. *Cell. Mol. Life Sci.* **64**, 2564–2574 (2007).
16. Cogdell, R. J. *et al.* The structure and function of the LH2 (B800–850) complex from the purple photosynthetic bacterium *Rhodospseudomonas acidophila* strain 10050. *Prog. Biophys. Mol. Biol.* **68**, 1–27 (1997).
17. Stock, D., Leslie, A. G. & Walker, J. E. Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1700–1705 (1999).
18. Sukharev, S., Durell, S. R. & Guy, H. R. Structural models of the MscL gating mechanism. *Biophys. J.* **81**, 917–936 (2001).
19. Strop, P., Bass, R. & Rees, D. C. Prokaryotic mechanosensitive channels. *Adv. Protein Chem.* **63**, 177–209 (2003).
20. Beckstein, O. & Sansom, M. S. P. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Phys. Biol.* **1**, 42–52 (2004).
21. Cruickshank, C. C., Minchin, R. F., Le Dain, A. C. & Martinac, B. Estimation of the pore size of the large-conductance mechanosensitive ion channel of *Escherichia coli*. *Biophys. J.* **73**, 1925–1931 (1997).
22. Ajouz, B., Berrier, C., Garrigues, A., Besnard, M. & Ghazi, A. Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells. *J. Biol. Chem.* **273**, 26670–26674 (1998).
23. van den Bogaart, G., Krasnikov, V. & Poolman, B. Dual-color fluorescence-burst analysis to probe protein efflux through the mechanosensitive channel MscL. *Biophys. J.* **92**, 1233–1240 (2007).
24. Yoshimura, K., Usukura, J. & Sokabe, M. Gating-associated conformational changes in the mechanosensitive channel MscL. *Proc. Natl Acad. Sci. USA* **105**, 4033–4038 (2008).
25. Strop, P. & Brunger, A. T. Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci.* **14**, 2207–2211 (2005).
26. Spencer, R. H. & Rees, D. C. The α -helix and the organization and gating of channels. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 207–233 (2002).
27. Wang, W. *et al.* The structure of an open form of an *E. coli* mechanosensitive channel at 3.45 Å resolution. *Science* **321**, 1179–1183 (2008).
28. Maurer, J. A., Elmore, D. E., Lester, H. A. & Dougherty, D. A. Comparing and contrasting *Escherichia coli* and *Mycobacterium tuberculosis* mechanosensitive channels (MscL). *J. Biol. Chem.* **275**, 22238–22244 (2000).
29. Tsai, I. J. *et al.* The role of the periplasmic loop residue glutamine 65 for MscL mechanosensitivity. *Eur. Biophys. J.* **34**, 403–412 (2005).
30. Ajouz, B., Berrier, C., Besnard, M., Martinac, B. & Ghazi, A. Contributions of the different extramembraneous domains of the mechanosensitive ion channel MscL to its response to membrane tension. *J. Biol. Chem.* **275**, 1015–1022 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Shih for early cloning work, T. Walton for biochemical analysis, A. Lee for initial efforts on the MscL project, J. Choe for expression and purification protocols, O. Lewinson for manuscript reading, Y. Poon and J. Lai for treating MJF465 with λ DE3 and the osmotic down-shock assay protocol, J. Kaiser for suggestions on structure refinement, R. Phillips, E. Haswell and P. Pal for discussions, P. Blount for the MJF465 strain, and the staff at SSRL, the Advanced Light Source (ALS) and the Advanced Photon Source (APS) for technical support with crystal screening and data collection. We would like to acknowledge the Gordon and Betty Moore Foundation for support of the Molecular Observatory at Caltech. Operations at SSRL, ALS and APS are supported by the US Department of Energy and NIH. This work was supported in part by grants from the Howard Hughes Medical Institute and the National Institutes of Health (GM084211). C.S.G. was supported in part by postdoctoral fellowships from the National Institutes of Health and the Beckman Foundation. D.C.R. is an Investigator in the Howard Hughes Medical Institute.

Author Contributions Z.L. designed and performed the experiments in molecular biology, biochemistry, crystallography and the structure analysis. Z.L. and C.S.G. conducted the down-shock assays. C.S.G. was responsible for the protein reconstitution and electrophysiology. D.C.R. coordinated the project and contributed to the structure analysis. The manuscript was written by Z.L., C.S.G. and D.C.R.

Author Information The atomic coordinates and structure factors have been deposited at the Protein Data Bank with accession numbers 3HZQ. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.C.R. (dcrees@caltech.edu).

METHODS

Cloning and protein expression. The gene encoding SaMscL was cloned from genomic DNA by PCR and inserted into a pET15b vector between NdeI and BamHI sites. A stop codon at Glu 95 was introduced using the Quikchange II kit (Stratagene) to create the SaMscL(CA26) gene. The plasmid was transformed into BL21-Gold (DE3)-competent cells (Stratagene). A single colony was used to inoculate 200 ml Terrific Broth media with 1% glucose (TB glucose), and the culture was grown overnight at 37 °C with shaking at 225 r.p.m. Two litres of TB glucose media were inoculated with 40 ml of the overnight culture and shaken at 37 °C until the absorbance at 600 nm reached 2.0. Expression was induced with 1 mM IPTG at 30 °C for 2 h, after which the cells were spun down at 6,000g and stored at -80 °C.

Protein purification and crystallization. Ten grams of cell pellet were resuspended in 100 ml lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 44 mM LDAO, pH 8.0). The mixture was sonicated for 4 min in an ice water bath (1 s on and 3 s off) and centrifuged at 17,000 r.p.m. (JA-17, Beckman) for 40 min at 4 °C. The supernatant was applied to a 2-ml Ni-NTA (Qiagen) column by gravity and washed with 10 ml high salt buffer (20 mM Tris-HCl, 300 mM NaCl, 30 mM imidazole, 5 mM LDAO, pH 7.5) and 10 ml low imidazole buffer (20 mM Tris-HCl, 20 mM NaCl, 75 mM imidazole, 5 mM LDAO, pH 7.5). Protein was eluted in 6 ml high imidazole buffer (20 mM Tris-HCl, 20 mM NaCl, 300 mM imidazole, 5 mM LDAO, pH 7.5), concentrated to ~15 mg ml⁻¹ in 30 kDa cutoff concentrators (Amicon Ultra-4, Millipore), and loaded onto a Superdex 200 10/30 HR column (GE Healthcare) equilibrated in 10 mM Tris-HCl, 150 mM NaCl, 2 mM LDAO, pH 7.5. The major peak fraction eluting at ~14.3 ml was collected, concentrated to 15–20 mg ml⁻¹ protein and used directly for crystallization or frozen in liquid nitrogen for storage at -80 °C.

Crystallization hits for the SaMscL(CA26) protein were originally found with the MemGold screen (Molecular Dimensions). Subsequent optimization yielded crystals that matured in a month to a maximum dimension of ~0.5 mm under a condition with 24–30% PEG400, 100 mM Tris-HCl (pH 7.0) and 150 mM Na₂SO₄. The addition of 3.0 mM decyl-β-D-maltoside (DM) to the protein sample improved the crystal size and reproducibility of crystal growth. The diffraction limit was consistently improved from 7 Å to 4.7–4.2 Å resolution, when crystals grown in sulphate were crosslinked with 25% glutaraldehyde through vapour diffusion at 4 °C for 1–2 h according to ref. 31 and then soaked overnight in 10 mM Na₂WO₄ or Na₂MoO₄. For phasing and establishing the sequence register, a series of cysteine mutants of SaMscL(CA26) at A9, S32, A87, S57 and S69 were made and crystallized. One crystal of S32C diffracted to 3.8 Å resolution after soaking in a solution of 32% PEG400, 100 mM Tris-HCl (pH 7.0), 20 mM Na₄P₂O₇, 200 mM LiCl, 1.8 mM DM, 4.3 mM LDAO and 1.0 mM CH₃HgCl for 4 days. Although this S32C crystal was soaked with CH₃HgCl, the mercury for some reason did not bind to the cysteine residue under this soaking condition. The data set collected from this crystal was used to refine the final structure model reported here.

Data collection and structure determination. Native 1 data set and native 2 data set (S32C) were collected on BL11-1 and BL12-2 of the Stanford Synchrotron Radiation Lightsource (SSRL), respectively. The Na₂WO₄ anomalous data set was collected on a laboratory RaxisIV++ system (Rigaku). The A87C + CH₃HgCl anomalous data set was collected on SSRL BL12-2.

The initial structure was solved by the method of multiple isomorphous replacement with anomalous scattering. The native 1 data set was collected from a molybdate-soaked crystal, while tungstate-soaked and A87C+Hg crystals provided two derivative data sets with strong anomalous signals. The heavy atom positions were determined with ShelxCDE³², and SigmaA³³ in the CCP4 suite³⁴ was used to combine the initial phases output by ShelxCDE, yielding a figure-of-merit (FOM) of 0.495 to an effective resolution of 4.7 Å. After density modification by the DM program³⁵, the resolution for the phase set was extended to 4.2 Å and the FOM was improved to 0.715. The program O³⁶ was used for model building and adjusting. The initial model was built using the 4.2 Å resolution experimental map (Supplementary Fig. 2a) with anomalous difference Fourier maps (Supplementary Fig. 2b) calculated for the S32C, S69C and A87C crystals (all with Hg bound) to establish the sequence register. The model was first refined against 4.2 Å resolution data with CNS program version 1.2 (ref. 37) to $R_{\text{work}}/R_{\text{free}} = 0.335/0.359$. When the 3.8 Å resolution data became available, the model was further refined against this larger data set. The application of a -160 Å² B-factor sharpening in the map calculation was helpful to show side-chain electron density. After a few rounds of model adjustment and refinement, $R_{\text{work}}/R_{\text{free}}$ converged to 0.291/0.312. The final $2F_o - F_c$ electron density is well defined in the transmembrane helices (Supplementary Fig. 2c, d) and the periplasmic loop (supplementary Fig. 2c), which allowed continuous tracing of the polypeptide from Gly 12 to Glu 94. The amino terminus (1–11) has relatively weak density and is modelled as an irregular structure instead of a short helix as

in MtMscL. Data collection and structure refinement statistics are summarized in Supplementary Table 1.

Structure analysis and model construction. Procheck³⁸ Ramachandran plot statistics for the final structure are 70.6%, 27.1%, 2.4% and 0.0% for the most favoured, additional allowed, generously allowed and outlier categories, respectively. The sequence alignment in Fig. 2e was generated with the programs ClustalW³⁹ and ESPrpt⁴⁰. Structural images were produced by PyMOL⁴¹. The HOLE program⁴² was used to evaluate the pore radius profile from structural models. The periplasmic surface area of the SaMscL(CA26) tetramer was approximated as an octagon enclosed by the α-carbon atoms of Gly 48 and Gly 60, whereas that of MtMscL resembles a circular area enclosed by the α-carbon atoms of Val 48. The structure of MtMscL in the closed state (Protein Data Bank code 2OAR), an intermediate model of EcMscL in dimyristoyl-phosphatidylcholine liposomes (1KYL) and the lysophosphatidylcholine stabilized open-state structure of EcMscL (1KYM) were downloaded from the Protein Data Bank. The open state model of Sukharev and Guy⁸ (SG model) was downloaded from the Nature website (http://www.nature.com/nature/journal/v409/n6821/extref/409720a0_S3.txt). The inter-TM1 angle was calculated by PROMOTIF⁴³ or directly measured in O³⁶. The inter-TM1 angles (α) for 2OAR (closed state), 1KYL, 1KYM and the SG model were determined to be -42°, -43°, -55° and -67°, respectively. The earlier observation that the crossing angles of the pore-lining helices are similar in the pentameric MscL and tetrameric KcsA⁶ indicates that it is feasible to apply the crossing angle data retrieved from a pentameric model to a tetrameric one. To generate the models for the first-step pivoting motion, the TM1–TM2' pair of the expanded-state structure was aligned to the Z axis and then pivoted clockwise about the X axis on a point around Val 21 to increase the crossing angle α from -42° to -68° in -2° steps. A program based on the equations reported in ref. 26, which describe the relationship among the TM1–TM1' crossing angle α, the tilt angle η of TM1 and the minimum pore radius R, was used for model building in the first step. The intermediate model generated at α = -62° aligns well with the observed structure of SaMscL(CA26). To make the second-step pivoting models, the model at the end point (with α = -68°) of the first step was rotated anticlockwise in 2° steps about an axis, passing through the α-carbon atom of Gly 48 and parallel to the Z axis, using the program PDBSET in CCP4³⁴.

Crosslinking. Protein was buffer exchanged into 100 mM MES (pH 5.0), 150 mM NaCl, 2 mM LDAO in a 30-kD cutoff concentrator. Reactions with 0.1–1.0 mg ml⁻¹ protein and 1–2 mM N,N-dicyclohexylcarbodiimide (DCCD) were incubated for 30–60 min at room temperature. An equal volume of 2× loading buffer (100 mM Tris-HCl, 4% SDS, 20% glycerol, 0.04% bromophenol blue, 200 mM β-mercaptoethanol, pH 6.8) was added to quench the reaction. The crosslinking products were electrophoretically analysed by 12% SDS-polyacrylamide gel electrophoresis and stained with Coomassie brilliant blue R250. At higher DCCD/protein ratios (Supplementary Fig. 3 lane b and lane d), a tetramer band dominated the crosslinking reaction of both CA26 and full-length SaMscL. Weak bands above the dominant species are probably generated by inter-tetramer crosslinking. Crosslinking with glutaraldehyde gave similar results as DCCD.

In vivo functional assay. MJF465 (*mscL*⁻, *mscS*⁻, *mscK*⁻)¹² cells treated with λDE3 lysogenization kit (Novagen) were transformed with pET15b plasmids containing either EcMscL (positive control), SaMscL, SaMscL(CA26), or the empty vector (negative control). The transformants were grown on LB-ampicillin (Amp) plates overnight. The down-shock assay protocol was modified from refs 12, 44. In brief, cells were grown and induced with 1 mM IPTG in LB medium with 500 mM NaCl (LB 500) and diluted 1:500 into sterile double-distilled H₂O (down-shock) or LB 500 (mock shock). The cells were recovered at 37 °C for 20 min and 20 μl of the shocked/mock-shocked culture was combined with 80 μl LB 500, spread onto LB-ampicillin plates and incubated overnight at 37 °C. At least four replicates were performed for each shock/mock shock experiment. The mock shock sets always gave a high percentage of surviving colonies for all constructs, including the empty vector.

Protein reconstitution and electrophysiology. Protein reconstitution, vesicle formation and electrophysiology methods were modified from refs 45 and 46. For reconstitution, 1–18 μg of protein in PBS + 0.05% DDM (w:v) was added to 200 μg of azolectin (Sigma) suspension in PBS. The protein/lipid/detergent mixture was incubated for 1 h at room temperature and an additional 3 h with 60 mg of Bio-Beads (Biorad) to remove DDM. After bead removal, the mixture was ultracentrifuged (160,000g, 50 min, 4 °C). The resulting pellet was washed once in DR buffer (200 mM KCl, 10 mM HEPES, pH 7.4) and resuspended in 40 μl of the same buffer. Ten microlitres of suspension was spotted onto the underside of a clean screw cap of the EasyXtal CrystalSupport plate (Qiagen) and dried overnight in a vacuum desiccator, depositing a thin lipid film. The film was wetted with 10 μl of DR buffer, equilibrated against 500 μl reservoir of DR buffer, and rehydrated overnight creating a mixture of uni- and multilamellar giant

vesicles. Two microlitres of this mixture was then loaded into a patch clamp recording chamber for analysis.

Unilamellar vesicles were visually identified and excised using patch pipettes with 3–9 M Ω resistances. Bath and pipette solutions were 200 mM KCl, 40 mM MgCl₂, 5 mM HEPES, pH 7.2. An HSPC-1 pressure clamp (ALA Scientific Instruments) was used to apply pressure jumps and ramps. The patches were voltage-clamped at +60 mV and current was recorded using an Axopatch 200b amplifier (Molecular Devices) and a 10 kHz sampling frequency. Single channel amplitudes were analysed using Clampfit (Molecular Devices).

31. Lusty, C. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J. Appl. Cryst.* **32**, 106–112 (1999).
32. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
33. Read, R. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38 (1994).
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
37. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
38. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
39. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
40. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
41. DeLano, W. L. *The PyMOL User's Manual* (Delano Scientific, 2002).
42. Smart, O. S., Neduvilil, J. G., Wang, X. N., Wallace, B. A. & Sansom, M. S. P. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph. Model.* **14**, 354–360 (1996).
43. Hutchinson, E. G. & Thornton, J. M. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).
44. Iscla, I., Wray, R. & Blount, P. On the structure of the N-terminal domain of the MscL channel: helical bundle or membrane interface. *Biophys. J.* **95**, 2283–2291 (2008).
45. Häse, C. C., Le Dain, A. C. & Martinac, B. Purification and functional reconstitution of the recombinant large mechanosensitive ion channel (MscL) of *Escherichia coli*. *J. Biol. Chem.* **270**, 18329–18334 (1995).
46. Sukharev, S. I., Martinac, B., Arshavsky, V. Y. & Kung, C. Two types of mechanosensitive channels in the *Escherichia coli* cell envelope: solubilization and functional reconstitution. *Biophys. J.* **65**, 177–183 (1993).

Direct observation of the binding state of the kinesin head to the microtubule

Nicholas R. Guydosh¹ & Steven M. Block^{2,3}

The dimeric motor protein kinesin-1 converts chemical energy from ATP hydrolysis into mechanical work used to transport cargo along microtubules^{1,2}. Cargo attached to the kinesin stalk moves processively in 8-nm increments³ as its twin motor domains (heads) carry out an asymmetric, 'hand-over-hand' walk^{4–7}. The extent of individual head interactions with the microtubule during stepping, however, remains controversial^{4,8–14}. A major experimental limitation has been the lack of a means to monitor the attachment of an individual head to the microtubule during movement, necessitating indirect approaches. Here we report the development of a single-molecule assay that can directly report head binding in a walking kinesin molecule, and show that only a single head is bound to the microtubule between steps at low ATP concentrations. A bead was linked to one of the two kinesin heads by means of a short DNA tether and used to apply rapidly alternating hindering and assisting loads with an optical trap. The time-dependent difference between forwards and backwards displacements of the bead alternated between two discrete values during stepping, corresponding to those intervals when the linked head adopted a bound or an unbound state. The linked head could only rebind the microtubule once ATP had become bound to its partner head.

Optical-trapping assays for kinesin typically involve tracking and loading a bead attached to the carboxy terminus of the common stalk of the protein³ in place of its natural cargo. Because beads attached to the stalk tend to report the position of the molecule as a whole, the displacements generated by the two separate heads remain unresolved. Single-molecule fluorescence techniques have successfully resolved motions of the heads by tagging these individually with fluorophores, but such approaches have suffered from limited spatiotemporal resolution and lack the ability to apply controlled loads, which can be used to probe binding between the heads and the microtubule^{4,8,9,14}.

We overcame these limitations by linking an optically trapped bead directly to one of the two heads, rather than to the stalk (Fig. 1a), without disabling the motor function of the molecule. One end of a double-stranded DNA oligomer (70 base pairs) was attached to an engineered cysteine residue (N62C) on the motor domain of an otherwise 'cysteine-light' dimer, and its opposite end was linked to a streptavidin-coated bead. Because this short oligomer subtends less than half a persistence length, the tether is stiff and transmits motions of the head directly to the bead. Control experiments in which the bead was attached instead to the stalk of an oligomer-linked construct showed that the motor still stepped in 8-nm increments (Supplementary Fig. 1). With the bead linked to one of the two heads, but unloaded, motors travelled an average distance of 980 ± 80 nm (mean \pm s.e.m.) before detaching (Supplementary Fig. 2), displaying a processivity equivalent to that of unlabelled kinesin¹⁵. Notably, the unloaded speeds of bead-linked

motors under these conditions averaged 581 ± 9 nm s⁻¹ (mean \pm s.e.m.), which is comparable to unloaded speeds recorded for wild-type motors (Supplementary Fig. 3). When head-linked beads were subjected to hindering loads under force-clamped conditions, we observed steps of 16.3 ± 0.1 nm towards the (+) end of the microtubule (Fig. 1b and Supplementary Fig. 4a), exactly as expected for individual heads executing a hand-over-hand walk along protofilaments with an 8.2-nm tubulin-dimer spacing. The stall force was halved when loads were applied to a single head instead of the stalk (falling from ~ 6 pN to ~ 3 pN)¹⁵, which is consistent with the energetic efficiency of the motor being unaffected by the head linkage, because half the load is moved through twice the distance per step (Supplementary Fig. 3).

When head-linked beads were subjected to assisting loads, we observed that the abrupt, 16-nm stepwise advances were generally composed of an 'overshoot' motion (~ 23 nm forwards) followed by a 'recovery' motion (~ 7 nm backwards) (Fig. 1c and Supplementary Fig. 4b). The durations of dwell intervals following the overshoot and recovery motions were sensitive to the ATP concentration, and both these intervals increased when the ATP concentration was reduced (Fig. 1d and Supplementary Fig. 5). This property indicates that the motor must bind ATP during both overshoot and recovery intervals. We also found that the positional variance of a head-linked bead was higher during overshoots than during recoveries, suggesting that additional compliance is introduced into the linkage between the bead and the microtubule during overshoots (Supplemental Fig. 6). Because of their small amplitudes, a fraction of recovery motions were missed by our step-finding procedure, causing some pairs of overshoot and recovery motions to appear as a single 16-nm transition.

The motions of head-linked beads under assisting loads may be explained by the following model (Fig. 1e), which is consistent with the established biochemical reaction cycle for kinesin. The binding of ATP to the microtubule-bound, linked head (state 1) triggers its unbound partner head to advance and rebind to the microtubule (state 2). (ATP binding and head motions represent separate steps, but are shown concomitantly for simplicity.) The linked head then releases the microtubule and moves ahead of its bound partner (state 3). In this case, the application of an assisting load not only helps to pull this head forwards, but also supplies a small torque, generated between the neck linker domain (joining the head to the stalk) and the point of attachment of the DNA. To relieve torque, the linked head rotates to adopt a backwards-facing orientation, free of the microtubule, during overshoot dwells. The neck linker region consists of a single polypeptide chain (~ 13 residues) and can act as a free swivel. Subsequent binding of ATP to the microtubule-bound head allows the free head to rebind to the microtubule in its proper orientation and release ADP (state 4). Finally, the rear head hydrolyzes ATP, releases P_i (inorganic phosphate), and unbinds from the microtubule to regenerate the starting configuration (state 1), but

¹Biophysics Program, ²Department of Biology, ³Department of Applied Physics, Stanford University, Stanford, California 94305, USA.

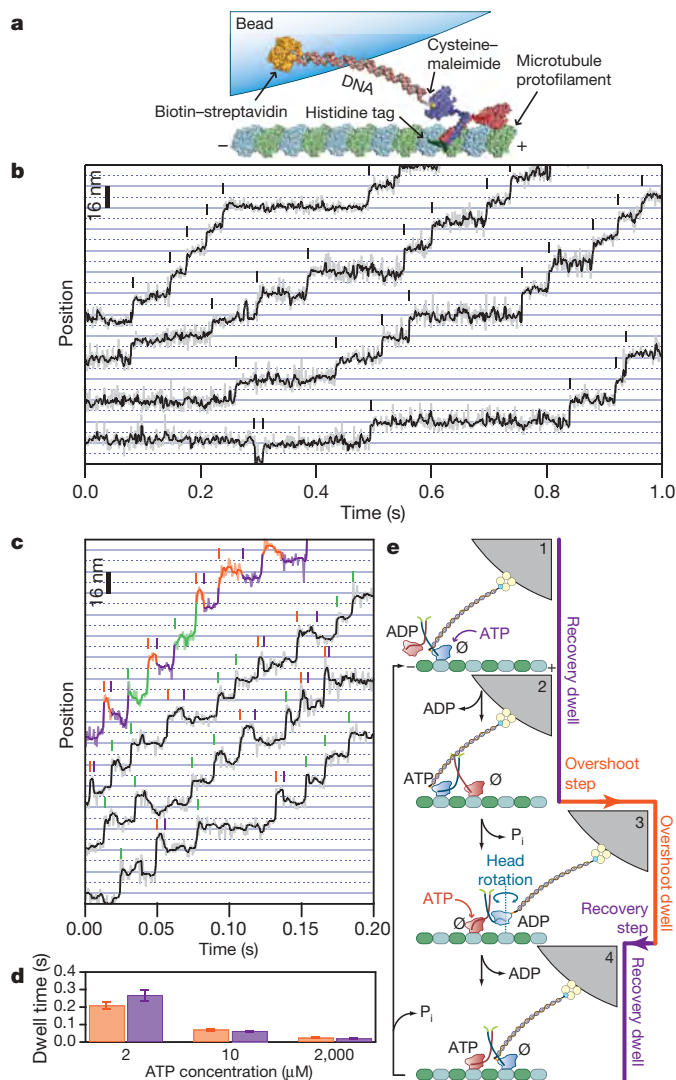


Figure 1 | Single-molecule observations of kinesin head motion.

a, Illustration of the experimental geometry (to scale)²⁹. **b**, Four representative displacement records of head-linked beads under force-clamped conditions (1.7 pN hindering load; 2 mM ATP). Lighter traces are unfiltered; darker traces are median-filtered. The records are offset vertically for clarity and stepwise transitions (~ 16 nm) are indicated by vertical dashes. **c**, Same as **b**, but under assisting load (1.7 pN). Overshoot (orange) and recovery (purple) transitions are indicated by vertical dashes. Corresponding dwell intervals are coloured in the uppermost record. For separate overshoot and recovery intervals that were unresolved, the combined dwell interval and associated transition are coloured green. **d**, Dependence of overshoot (orange) and recovery (purple) dwell times on ATP concentration (mean \pm s.e.m.). Uncertainties were computed by bootstrapping. Load, 1.7 pN. **e**, Model for kinesin motions under assisting load, cycling through two steps. Heads are labelled according to the nucleotide bound (ADP, ATP or \emptyset (no nucleotide)).

with the molecule advanced by 16.4 nm (two steps). The existence of an overshoot implies that kinesin adopts a one-head-bound (1-HB) 'ATP waiting state' between successive 8-nm advancements of its stalk. Furthermore, the dependence of the overshoot dwell time on ATP suggests that the unbound head cannot be induced to rebind to the microtubule simply by coaxing it in front of its partner with the optical trap: some additional requirement must be met.

If kinesin adopts a 1-HB state between processive advances of its stalk, as implied in the preceding discussion, then it should be possible to swing the free, linked head backwards and forwards with respect to its bound partner by reversing the direction of load. To test this prediction, we modified our force-clamp apparatus to switch the sign of the constant load between hindering and assisting every 14 ms

as kinesin moved in the presence of 2 μ M ATP (this rate was chosen to be ~ 10 -fold faster than the kinesin stepping rate at this ATP concentration, but ~ 10 -fold slower than the update time of the force clamp). Separate records were collected for displacements under each sign of load, and from these we computed the running difference between records. The value of the running difference is expected to increase whenever the linked head releases the microtubule and is alternately pulled in front of and behind its bound partner. Rebinding of the linked head to the microtubule will eliminate any extra slack introduced in the linkage between the bead and microtubule as a result of head unbinding, and will restore the running difference to its lower value. Accordingly, abrupt jumps in the running-difference record signify that the linked head adopts both bound and unbound states during the kinetic cycle, consistent with a 1-HB ATP waiting state. By contrast, the absence of such jumps would indicate that the linked head spends the majority of its reaction cycle bound to the microtubule, corresponding to a two-heads-bound (2-HB) ATP waiting state.

When alternating loads (± 1.7 pN) were applied to a moving kinesin molecule, we observed clear transitions in the difference record, averaging 23 nm in amplitude, marking the binding transitions of the linked head (Fig. 2). (Similar transitions were observed when the magnitude of the load was reduced to its lowest practical value, 0.4 pN; Supplementary Fig. 7). The presence of transitions is the signature of a 1-HB state. On the basis of an analysis of dwell times in the running-difference record, we estimated that the linked head spends $\sim 93\%$ of its stepping cycle in an unbound state (Supplementary Discussion). Stepwise increases in the difference record coincided with overshoot motions under assisting loads, marking events where the head released the microtubule, was pulled forward and then rotated about its neck linker. These unbinding events also corresponded to tiny backwards motions under hindering loads, which probably results from a small increase in system compliance, as the linked head dissociates from the microtubule. Stepwise decreases in the difference record coincided with recovery motions under assisting loads, marking events where the head bound to the microtubule in front of its partner. These transitions corresponded to large forwards motions of the bead under hindering loads.

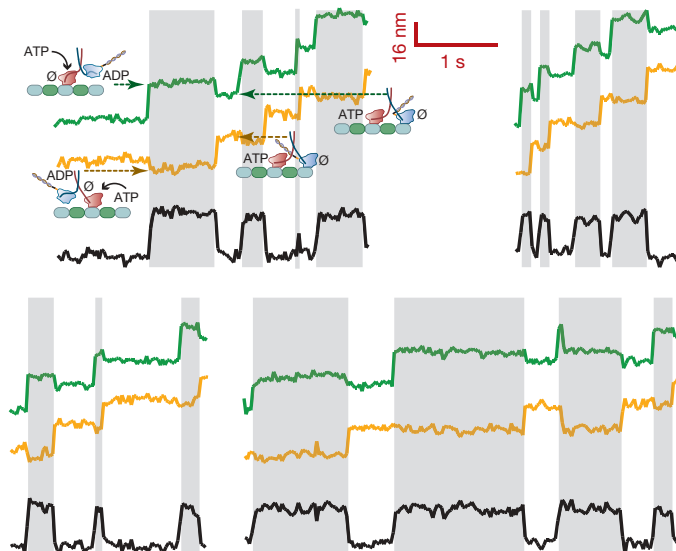


Figure 2 | Motion of a bead linked to one of the kinesin heads under alternating loads. Four records of bead position, recorded during intervals of alternating assisting (green) and hindering (orange) loads. The running difference is also shown (black). The inset sketches show conformations assigned to the molecule during each plateau. Regions with shading highlight intervals in which the linked head is unbound. Load, ± 1.7 pN; ATP concentration, 2 μ M.

To assess better whether the linked head remained free of the microtubule while its partner waited to bind ATP, we plotted the size of the transition under assisting load against the corresponding transition under hindering load (Fig. 3, central panel). Bead displacements that would normally be too small to resolve independently could be included in this analysis because the larger, readily detectable transitions under the opposite loading condition served as event markers. As expected, displacements fell into one of two populations, corresponding to unbinding (quadrant II) or rebinding (quadrant IV) events. We did not observe a population in which the corresponding displacements under both assisting and hindering loads were comparable, which would have signified a 2-HB ATP waiting state: just 6% of all transitions (mostly events in quadrant I) were in this category, and probably resulted from rare occasions when the unbound lifetime of the linked head was less than the time for load reversal of the force clamp. The small backwards motion under hindering load averaged -1.7 ± 0.2 nm. When added to the forward motion, 18.2 ± 0.2 nm, this yields a net displacement of 16.5 ± 0.3 nm, which is statistically indistinguishable from the expected advance of 16.4 nm (twice the tubulin-dimer spacing). Similarly, the sum of the overshoot (21.1 ± 0.3 nm) and recovery (-4.8 ± 0.3 nm) displacements yielded a net advance of 16.3 ± 0.4 nm. These results provide a strong consistency check on our interpretations and calibrations.

To exclude the possibility that the load applied by the optical trap might somehow pull the head free from the microtubule, we computed an estimate of the increase in head detachment rate as a consequence of load. Taking 2 nm as a (generous) estimate of the distance between the bound and transition states along the reaction coordinate for head unbinding¹⁶, the mechanical work available to detach a bound head would be less than $1.7 \text{ pN} \times 2 \text{ nm} = 3.4 \text{ pN nm}$. Because this value is less than the thermal energy ($k_B T \approx 4 \text{ pN nm}$, where k_B denotes Boltzmann's constant and T the temperature), any increase in detachment rate would be limited to a factor of two to three, and would be too small to account for our observations. This conclusion is reinforced by the persistence of transitions in the difference record when the load was

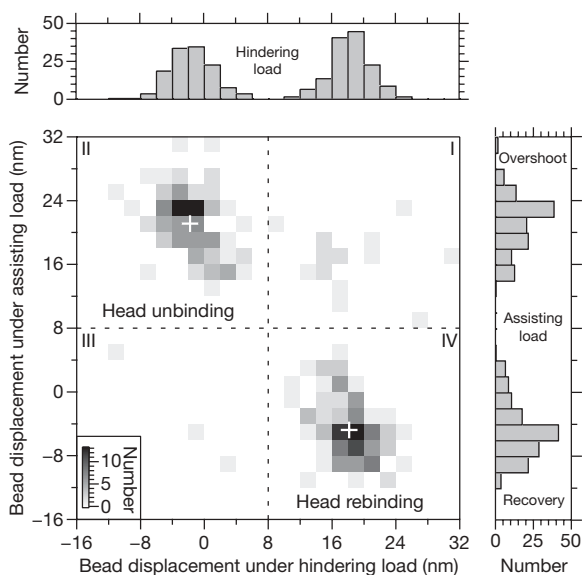


Figure 3 | Histograms of bead displacements under alternating load. The size of a displacement under assisting load was plotted against the size of the corresponding displacement under hindering load, and a two-dimensional histogram (central panel) was created for the data set. Two populations are evident, arising from unbinding (quadrant II) and rebinding (quadrant IV) of the linked head, as expected for a 1-HB ATP waiting state. A 2-HB state would populate points on the diagonal. White crosses mark the centroids of the two populations. Data from these quadrants are projected onto one-dimensional histograms (top and side panels). $N = 292$ (eight points landed outside the range plotted; not displayed) from 13 motors. Same load and ATP concentration as in Fig. 2.

reduced to 0.4 pN, which can only supply 3.2 pN nm of energy over an 8-nm step (Supplementary Fig. 7).

Our data support a model for the reaction cycle (Fig. 4) in which kinesin binds ATP and takes its step from a 1-HB state. Because low ATP concentrations or high loads can cause the transition from this state to become rate limiting, it is the longest-lived state of the cycle (and therefore the 'waiting state') under the conditions explored here. The unbound head is mobile, and can be readily pulled about its bound partner and even rotated with an optical trap. Importantly, our data do not support models for kinesin where both heads bind to the microtubule until ATP binding triggers release of the rear head^{4,8}, which postulate a long-lived, 2-HB state at low and high ATP concentrations. Our data also fail to support models for kinesin where one head is parked against the forward face of its microtubule-bound partner in the ATP waiting state¹⁰. However, our data may be reconciled with a recent study⁹ which concluded that one head remains poised behind its microtubule-bound partner head during the ATP waiting state, and either unbound, or transiently (weakly) bound, to the microtubule. In that study, estimates of the distance between the kinesin heads were obtained from fluorescence resonance energy transfer measurements recorded at 100 Hz and filtered using a 100-ms window, so the proposed structure probably corresponds to an average over an ensemble of positions. Our data clearly indicate that the unbound head can move readily around its equilibrium location.

The existence of a 1-HB state implies some form of mechanochemical gating to prevent the unbound head from rebinding the microtubule and releasing ADP (Fig. 4, S1 and S2). We propose that unfavourable strains developed in (competing) neck-linker conformations that correspond to off-pathway, 2-HB states constitute the mechanical coupling required for gating. According to this proposal, both heads will bind the microtubule only when the neck linker of the rear head is docked and the neck linker of the front head is undocked. The neck linkers are postulated to be too short, or too constrained, to adopt alternative conformations that might produce any alternative 2-HB state (Fig. 4, S1–S3). Cryo-electron microscopy structures of microtubule-bound kinesin^{17,18} show that the conformation of the kinesin neck linker is tightly coupled to the state of the nucleotide bound to the head.

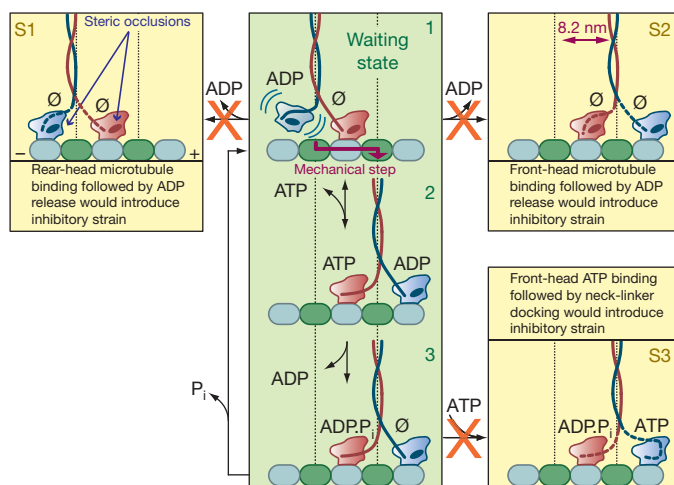


Figure 4 | A model for strain-based mechanochemical gating. A simplified kinesin cycle with three states (green panel). In the ATP waiting state (state 1), the rear kinesin head is unbound and mobile. Configurations postulated to produce inhibitory strain between the heads are indicated by showing neck linkers as dashed lines. This strain suppresses entry into off-pathway, 2-HB configurations (yellow panels; S1–S3). Strain is created when the neck linkers of either both kinesin heads (S3) or neither (S1, S2) dock to their catalytic cores. The 2-HB configuration is only allowed (states 2 and 3) when the neck linker of the rear head is docked (bound to ATP or ADP·P_i) and the neck linker of the front head is undocked (bound to ADP or Ø). Structural elements (dark ovals) inhibit neck-linker docking to a head bound to ADP or Ø.

When the head is either free of nucleotide or binds ADP, an occluding conformation of the switch-II helix prevents docking of the neck linker against the catalytic core¹⁹. By contrast, the motor core is tilted when bound to ATP or to analogues of ADP.P_i, allowing the neck linker to dock. Therefore, if both heads are to bind the microtubule simultaneously, the rear head must bind ATP (or ADP.P_i) while the front head holds ADP or no nucleotide. In support of this, recent work suggests that kinesin does not uniformly adopt a 2-HB configuration in the presence of pyrase⁹.

The inability of one head to bind the microtubule offers a natural explanation for the observation^{7,20–22} that the microtubule-stimulated release of ADP is inhibited until the microtubule-attached head binds ATP and docks its neck linker (Fig. 4, state 2). Strain produced by an unfavourable neck-linker conformation also explains the observation^{23–25} that ATP does not bind prematurely to the front, nucleotide-free head of a 2-HB kinesin molecule (Fig. 4, state 3). Any tight binding of ATP is disfavoured because it is coupled to neck-linker docking²⁶ and, therefore, to the generation of a strained configuration²⁷ in which both neck linkers are docked (Fig. 4, S3). We anticipate that the single-molecule techniques presented here will be applicable to the study of dynamic properties of other motors and macromolecules that undergo analogous conformational rearrangements.

METHODS SUMMARY

Labelling kinesin with DNA. We introduced mutation N62C into a 'cysteine-light' construct consisting of the first 401 residues of the *Drosophila melanogaster* heavy chain followed by a histidine tag. A 70-base-pair double-stranded DNA oligomer was synthesized with a primary amine at the 5' end of one strand and biotin at the 5' end of the other (Integrated DNA Technologies). We affixed these labels to the DNA using hydrocarbon linkers (6–12 carbon atoms long) intended to serve as swivels. DNA was activated with sulpho-SMCC (Pierce) before incubation with kinesin. Labelled kinesin molecules were purified by ion-exchange chromatography (MonoQ, GE Healthcare).

Force clamp. We performed trap calibration and measurements under constant load as previously described²⁸. Oscillatory loading was implemented in custom force-clamp software (LABVIEW 7.1) by reversing the sign of separation between the bead and trap every 14 ms; laser position was updated every 2 ms. To reduce settling time after a load reversal, the trapping beam was initially deflected to its approximate new position under reversed load before initiating feedback. Only data recorded after the final force-clamp update and before load reversal were included in analysis, to allow time for the feedback to equilibrate. **Step sizes.** We estimated average bead displacements from the arithmetic means of records over the appropriate intervals. Uncertainties (s.e.m.) were computed by bootstrapping.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 1 July 2009.

Published online 19 August 2009.

- Block, S. M. Kinesin motor mechanics: binding, stepping, tracking, gating, and limping. *Biophys. J.* **92**, 2986–2995 (2007).
- Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112**, 467–480 (2003).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303**, 676–678 (2004).
- Kaseda, K., Higuchi, H. & Hirose, K. Alternate fast and slow stepping of a heterodimeric kinesin molecule. *Nature Cell Biol.* **5**, 1079–1082 (2003).
- Asbury, C. L., Fehr, A. N. & Block, S. M. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science* **302**, 2130–2134 (2003).

- Hackney, D. D. Evidence for alternating head catalysis by kinesin during microtubule-stimulated ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **91**, 6865–6869 (1994).
- Yildiz, A., Tomishige, M., Gennerich, A. & Vale, R. D. Intramolecular strain coordinates kinesin stepping behavior along microtubules. *Cell* **134**, 1030–1041 (2008).
- Mori, T., Vale, R. D. & Tomishige, M. How kinesin waits between steps. *Nature* **450**, 750–754 (2007).
- Alonso, M. C. et al. An ATP gate controls tubulin binding by the tethered head of kinesin-1. *Science* **316**, 120–123 (2007).
- Hackney, D. D. The tethered motor domain of a kinesin-microtubule complex catalyzes reversible synthesis of bound ATP. *Proc. Natl Acad. Sci. USA* **102**, 18338–18343 (2005).
- Kawaguchi, K., Uemura, S. & Ishiwata, S. Equilibrium and transition between single- and double-headed binding of kinesin as revealed by single-molecule mechanics. *Biophys. J.* **84**, 1103–1113 (2003).
- Mandelkow, E. & Hoenger, A. Structures of kinesin and kinesin-microtubule interactions. *Curr. Opin. Cell Biol.* **11**, 34–44 (1999).
- Asenjo, A. B. & Sosa, H. A mobile kinesin-head intermediate during the ATP-waiting state. *Proc. Natl Acad. Sci. USA* **106**, 5657–5662 (2009).
- Block, S. M., Asbury, C. L., Shaevitz, J. W. & Lang, M. J. Probing the kinesin reaction cycle with a 2D optical force clamp. *Proc. Natl Acad. Sci. USA* **100**, 2351–2356 (2003).
- Taniguchi, Y., Nishiyama, M., Ishii, Y. & Yanagida, T. Entropy rectifies the Brownian steps of kinesin. *Nature Chem. Biol.* **1**, 342–347 (2005).
- Sindelar, C. V. & Downing, K. H. The beginning of kinesin's force-generating cycle visualized at 9-Å resolution. *J. Cell Biol.* **177**, 377–385 (2007).
- Kikkawa, M. & Hirokawa, N. High-resolution cryo-EM maps show the nucleotide binding pocket of KIF1A in open and closed conformations. *EMBO J.* **25**, 4187–4194 (2006).
- Hwang, W., Lang, M. J. & Karplus, M. Force generation in kinesin hinges on cover-neck bundle formation. *Structure* **16**, 62–71 (2008).
- Ma, Y.-Z. & Taylor, E. W. Interacting head mechanism of microtubule-kinesin ATPase. *J. Biol. Chem.* **272**, 724–730 (1997).
- Gilbert, S. P., Moyer, M. L. & Johnson, K. A. Alternating site mechanism of the kinesin ATPase. *Biochemistry* **37**, 792–799 (1998).
- Crevel, I., Carter, N., Schliwa, M. & Cross, R. Coupled chemical and mechanical reaction steps in a processive *Neurospora* kinesin. *EMBO J.* **18**, 5863–5872 (1999).
- Klumpp, L. M., Hoenger, A. & Gilbert, S. P. Kinesin's second step. *Proc. Natl Acad. Sci. USA* **101**, 3444–3449 (2004).
- Rosenfeld, S. S., Fordyce, P. M., Jefferson, G. M., King, P. H. & Block, S. M. Stepping and stretching. How kinesin uses internal strain to walk processively. *J. Biol. Chem.* **278**, 18550–18556 (2003).
- Guydosh, N. R. & Block, S. M. Backsteps induced by nucleotide analogs suggest the front head of kinesin is gated by strain. *Proc. Natl Acad. Sci. USA* **103**, 8054–8059 (2006).
- Rice, S. et al. A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
- Sablin, E. P. & Fletterick, R. J. Coordination between motor domains in processive kinesins. *J. Biol. Chem.* **279**, 15707–15710 (2004).
- Lang, M. J., Asbury, C. L., Shaevitz, J. W. & Block, S. M. An automated two-dimensional optical force clamp for single molecule studies. *Biophys. J.* **83**, 491–501 (2002).
- DeLano, W. L. PyMOL Molecular Viewer version 1.0r1 (<http://www.pymol.org>) (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Dunn, B. Choi and W. Hwang for advice on labelling kinesin; S. Gilbert for advice on expressing kinesin; and B. Gutiérrez-Medina, C. Sindelar, C. Perez and K. Frieda for comments on the manuscript. This work was supported by grant GM51453 from the US National Institutes of Health.

Author Contributions N.R.G. designed the project, expressed and labelled the protein, and collected and analysed data. N.R.G. and S.M.B. discussed the data and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M.B. (sblock@stanford.edu).

METHODS

Instrumentation and analysis. Data were acquired at 20 kHz, filtered at 1 kHz, then decimated and recorded at 2 kHz. We computed velocities by dividing the distance travelled by kinesin (the run length) by the time of a run, from unsmoothed records, and computed average velocities from the arithmetic mean of individual velocities weighted by run lengths. The locations of step transitions in records of bead motion were detected automatically³⁰ until (apparently) misidentified steps accounted for no more than one (alternating load) or two (sustained load) position(s) in the trial step staircase. We ignored those positions and recorded the remainder. The step detection algorithm was restricted to identifying steps separated by at least 2.5 ms. Rapid bead motions >32 nm were interpreted as dissociation events.

Under sustained hindering load, the force clamp was updated at 200 Hz and data were median-filtered before step detection, using a window size of 3 ms. We performed a second round of step detection by marking steps separated by >4.5 ms whenever the data, median-filtered using a 9-ms window, deviated by more than 8 nm from the original data, median-filtered using a 100-ms window. The initial and final dwells of all records were not included in the analysis. We computed the times of overshoot and recovery dwell intervals (Fig. 1d) only when the intervening recovery step was smaller than -6 nm. We therefore expect our experimental estimates of the overshoot and recovery dwell times to be slightly overestimated, because the separate detection of overshoot and recovery steps is difficult when the dwell times become very short. This 'missing-event problem' tends to suppress the trend of decreasing dwell times as the ATP concentration is increased (Fig. 1d), because a larger proportion of brief dwell times will be missed at high ATP concentrations than at low ATP concentrations. The actual dependence of dwell times on the ATP concentration may therefore be even stronger than that reported here (but does not change any of our conclusions).

Under alternating load, step positions detected under hindering and assisting loads were merged. Any steps separated by a single load reversal (14 ms) were taken as a single step transition at the midpoint between these. The final dwell interval and step transition in any given stepping record were excluded from the analysis.

In the absence of load, we used video frame capture (at 30 Hz) followed by image processing to track the centroid of a moving bead. Bead release from the microtubule was determined from the increase in its positional variance.

Single-molecule assay. We assembled plasma-cleaned, poly-lysine-coated coverslips were assembled into flow cells. Microtubules and 20 mg ml⁻¹ BSA blocking protein were added sequentially. Kinesin covalently linked to DNA was mixed with 0.44- μ m streptavidin beads (SpheroTech) and incubated overnight with ~6 mg ml⁻¹ BSA. Before measurements, the ATP level was adjusted to the final concentration and an oxygen-scavenging system was introduced (235 μ g ml⁻¹ glucose oxidase (Calbiochem), 42 μ g ml⁻¹ catalase (Roche) and 4.6 mg ml⁻¹ glucose). The experimental buffer consisted of 80 mM Pipes (pH 6.9), 50 mM potassium acetate, 4 mM MgCl₂, 2 mM DTT, 1 mM EGTA and 10 μ M Taxol.

We recorded data from samples in which only 10–35% of beads moved. Assisting loads greater than 1.7 pN often caused the DNA-linked kinesin irreversibly and abruptly to lose activity and then stick to the cover-glass surface. Such loads may trigger unfolding of the motor and were therefore avoided. In contrast, sustained hindering loads as high as 3 pN rarely incapacitated the motor. We attached control beads directly to the kinesin stalk by incubating kinesin with beads coated with anti-histidine antibody.

Kinesin purification and labelling. The gene for cysteine-light *Drosophila* kinesin-1 (ref. 31) was mutated at position 62 from asparagine to cysteine. We selected this residue as an attachment point because it is situated in a solvent-exposed loop on the back side of the motor domain, distal from regions known to be critical for motility, such as the neck linker and nucleotide-binding pocket. Two additional constructs were generated that successfully generated 16-nm steps when beads were linked to a single head, A128C and S181C (data not shown). The N62C construct was expressed in *Escherichia coli* by inducing cultures grown at 37 °C to relative absorbance of 0.4 (at 600 nm) with 100 μ M IPTG. We then incubated cultures at 20 °C overnight. Cell lysates were first purified with Ni-NTA affinity resin (Qiagen) and then by means of ion exchange (MonoQ, GE Healthcare) using a buffer gradient composed of 25 mM bis-tris propane (pH 7.4 at 4 °C), 2 mM MgCl₂, 1 mM EGTA, 5 mM DTT, 2 μ M ATP, and 100–1,000 mM NaCl. This procedure yielded ~1.5 mg dimeric kinesin per litre of bacterial culture. Kinesin fractions were confirmed by mass spectrometry and SDS-PAGE.

A solution of double-stranded DNA oligomers (100 μ M) and sulpho-SMCC (2 mM) was incubated for 1 h at 37 °C in 25 mM potassium phosphate buffer (pH 7.2) containing 100 mM NaCl. The sample was passed over a desalting column (Micro Bio-Spin 6, BioRad) three times to remove unreacted sulpho-SMCC. In preparation for the reaction with DNA, we twice passed kinesin over desalting columns (Zeba, Pierce) to remove DTT and exchange it into the same buffer as the activated DNA. A mixture of ~22 μ M dimeric kinesin and ~67 μ M DNA was incubated for 12 h at 4 °C. ATP was added to bring the concentration to ~50 μ M. We quenched the reaction with 1 mM reduced cysteine and purified it using ion-exchange chromatography, as described above. Labelled and unlabelled kinesin was readily distinguished by a gel shift in native PAGE. The labelling efficiency was <25%, ensuring that >93% of labelled kinesins were attached to a single DNA oligomer. We found that <10% of beads nominally attached to DNA-linked kinesin took 8-nm steps (rather than ~16-nm steps) under sustained hindering loads. We attributed these to beads carrying double motor linkages or to conjugation between the DNA and stalk residues, and excluded them from the analysis. The sequence of the amine-labelled strand of the 70-base-pair DNA oligomer was 5'-CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG-3'.

30. Kerssemakers, J. W. *et al.* Assembly dynamics of microtubules at molecular resolution. *Nature* **442**, 709–712 (2006).

31. Fehr, A. N. Gutiérrez-Medina, B. Asbury, C. L. & Block, S. M. On the origin of kinesin limping. *Biophys. J.* (in the press).



Visa problems continue to frustrate scientists seeking entry to the United States. **Karen Kaplan** investigates the roadblocks.

Scientists, postdocs and students planning to travel to the United States to work or study need two things before applying for a visa: time and patience.

Despite recent efforts by federal agencies to improve and accelerate the visa-application process — including adding staff and setting shorter waiting times — it still needs legislative and regulatory reform, say those who are familiar with the system. Many consider it to be a labyrinthine muddle of requirements and regulations. Delays of up to half a year are not uncommon, even with the processing improvements brought in to clear the backlog and speed procedures after the terrorist attacks of 11 September 2001 forced a visa clampdown.

The consequences of such a system are potentially much more damaging than inconveniencing a few researchers and students seeking to work or study temporarily in the country. In a January 2009 report, the US National Research Council called for a thorough revamp of existing visa regulations, saying that the current system is quashing international scientific collaborations (see *Nature* **457**, 752; 2009). Some researchers are worried that the delays might mean collaborators give up coming altogether (see 'Unwelcome guest', overleaf).

In June, 31 US academic and research associations asked the federal government to streamline the visa process for international

scientists, scholars and students (see *Nature* **459**, 1157; 2009). The US National Postdoctoral Association also called for regulatory changes earlier this year, including a new non-immigrant visa for postdocs.

In response, the US Department of State, which handles student and exchange visa applications, brought in a combination of new procedures and extra staff and resources to expedite the process and chip away at the backlog that has contributed to delays.

These procedures are specifically intended to address those applications that require special 'administrative processing' — which is triggered when the applicant's work is linked to sensitive technology, weaponry or military applications, or if the applicant is from a country identified as a "sponsor of terror" or as a nuclear-proliferation concern. Most of these applications are now expected to get through the system in a maximum of ten working days, according to David Donahue, director of the public-affairs office at the state department. He says that routine applications that don't require special security checks account for 97% of the annual total — which last year reached some 6.5 million — and should take well under 30 days. Students in particular are benefiting from the shorter processing times, says Donahue. "We are doing everything we can to ensure that students don't miss the beginning of their study programmes," he says.

Although the state department brought

in the changes at the end of May, by late July university administrators, immigration lawyers and others who work with visa applicants had noticed little difference. They advise that an applicant should still allow up to six months for visa approval, although they acknowledge that in many cases, the visa will be issued within two to four weeks.

"It may be a little slow, it may not be the easiest interview, but people are getting through it and getting here — provided they start their application paperwork in a timely fashion," says Roger Chalkley, senior associate dean for biomedical research education and training at Vanderbilt University School of Medicine in Nashville, Tennessee. The school admits about 75 non-US graduate students and 250 non-US postdocs every year. "There have been hiccups, and it took a while, but things are getting resolved," he says.

Processing problems

Administrators still warn of persistent delays for those scientists or students already in the United States who need a different visa from the one they currently have. That could include a graduating student who needs an exchange visa, or a postdoc who is offered a permanent job and thus needs an employment visa.

Such employment visas — generally H-1Bs for scientists — are handled by the US Department of Homeland Security (DHS)

UNWELCOME GUEST

Juhn-Jong Lin, a condensed-matter physicist and professor at the Institute of Physics at the National Chiao Tung University in Hsinchu, Taiwan, hasn't been to the United States for two years. And he doesn't plan to return any time soon.

Lin used to hand his passport to a travel agent in Taiwan who would apply on his behalf for a J-1 exchange visa. Without fail, he would receive a five-year visa, which made it easy for him to attend the American Physical Society's biannual conferences and to travel to and from the United States for any other work-related reason as

often as he liked within that period. But things became more difficult after the terrorist attacks of 11 September 2001, and worsened again about two years ago, Lin says.

"After 9/11, I had to apply in person — and I always got only a one-year visa," he says. Every year he had to complete a lengthy application, prepare a detailed CV, have a new specific-sized photo taken, undergo an in-person interview 80 kilometres away in Taipei and pay an application fee of about US\$150. Waiting times for visa approval stretched from a week or two before 11 September 2001

to about two months, and recently ballooned again to four months. "No one knows why," Lin says.

In 2005, Lin planned to attend the American Physical Society's March meeting as usual and applied for a visa about eight weeks in advance. The document was not issued in time for Lin to attend.

Lin's experience is not unusual, says immigration lawyer Paul Herzog, who works in Los Angeles, California. Lin's work as a physicist means his visa application is likely to be flagged for a Visa Mantis screen, one of several types of security checks required for

'administrative processing'. Herzog says applications singled out for such processing are known to cause problems. "Those checks can take weeks and months," he says, "especially if [the federal agencies reviewing the applications] are concerned about espionage or terrorism."

Lin, for his part, says he has given up. "I no longer want to try," he says of applying in the future for a US visa. Lin has now formed fruitful collaborations with physicists in Japan, China and Europe. "There is no longer much incentive to visit the United States," he says. **K. K.**

and have lengthier processing times than the student or exchange visa. Chris Rhatigan, a spokeswoman for US Citizenship and Immigration Services, part of the DHS, says H-1B approval is usually received in two months, possibly longer depending on the applicant. "We don't post processing times," she says. "Each application is reviewed on a case-by-case basis."

Applicants already in the United States on a J-1 exchange visa who can't support themselves financially while they wait for H-1B approval must return to their home country. However, the process can be fast-tracked if their institution pays a 'premium-processing' fee of US\$1,000, which cuts the wait from 2–4 months to 15 days or fewer. University administrators say, however, that such an expenditure cannot be made routinely, especially by a research-intensive institution that might employ dozens of non-US postdocs every year.

"It's a significant cost to the sponsoring institution," says Mary Anne Timmins, administrative director of biomedical postdoctoral programmes at the University of Pennsylvania in Philadelphia. Other visa charges on top of the premium-processing fee can push the total cost of bringing in one non-US postdoc to more than \$3,000, she says.

Those affected by delays or the requirement to return home often wonder whether they should even go through such an upsetting process, says Peter Palese, head of microbiology at the Mount Sinai School of Medicine in New York, who works with international postdocs. "Other countries make it easier," he says. "We are harming ourselves."

Some administrators and others say that despite improvements to the process for all types of special security checks — including those conducted under Visa Mantis, the check that most commonly affects scientists, science students and postdocs — wait times are still far longer than ten days. Under a Mantis check, federal agencies including the state department, the DHS, the FBI and

the CIA investigate an application for links to terrorism, espionage or illegal transfer of sensitive technology.

"Things aren't as bad as they were right after 9/11," says immigration lawyer Elizabeth Goss of legal firm Tocci, Goss and Lee in Boston, Massachusetts, "but there are still issues. And there's really no one you can talk to. That can be frustrating."

Heightened security

In 2003, special security checks took roughly 75 days to complete, according to independent federal investigators. After congressional pressure, the waiting time for Visa Mantis applicants was cut to about 15 days by the end of 2004. But in the past year, delays on Mantis applications lengthened again to between 4 and 12 months.

Although the Mantis checks make sense in principle, many say that in practice they clog up a process that should move far more freely. "The visa process should serve as a barrier to people with criminal or terrorist intent," says Vic Johnson, senior adviser for public policy at NAFSA, the Association of International Educators. "But it should

also be a gateway for people with the talent our economy and society requires." Before September 2001, Mantis clearances totalled several dozen a year. Last year, under Mantis, federal screeners reviewed some 55,000 applications. "If all these Mantis clearances are being approved, it must mean you're reviewing a lot that don't need review," says Johnson.

Still, officials say that further improvements are on the way. As well as the state department's efforts to speed up Visa Mantis processing, some US embassies have updated their websites to show the visa waiting times at consulates in every city.

An inter-agency group led by the DHS that meets to discuss visa issues convened most recently in July, and has met at the White House with the National Security Council to identify other areas of the system that need improvement. "We are always asking, 'Can we do better?'" Donahue says.

Legislative changes may also be coming. According to Amy Scott, assistant vice-president for federal relations at the Association of American Universities in Washington DC, a Senate bill could be introduced this month that would streamline the visa process for scientists and students. Although Scott says it is likely that the bill's passage will be delayed, she notes that President Barack Obama is calling for immigration reform by no later than next year.

Donahue and Rhatigan say their agencies recognize that streamlining the visa process allows for the free flow of scientists and their ideas into the United States — a positive outcome for all. "We will continue to listen to the science community and try to figure out ways to address their concerns," Donahue says. "There have been long waits for a variety of reasons. We have made a great effort to change that."

Karen Kaplan is assistant editor of Naturejobs.

See Editorial, page 12.



Physicist Juhn-Jong Lin has given up on US visas.

Fine-tuning the Universe

Touched by the hand of God.

Merrie Haskell

Titanium Sun Occluded, King of Earth (or in the common parlance of the day, the Greatest Equal Citizen), heard a court case in the 9.83923×10^8 second of his reign; and although Titanium Sun Occluded judged that programming a belief in creationism in the newborn was both illegal and illogical, the debate continued to rage for many years.

The King never spoke publicly on the subject again, once his judgement had been announced. But in private, Citizen Brilliant Cobalt, the King's favourite concubine, took up the subject again and again, so that the King could not rest easily with his decision.

"And I suppose you think you were a toaster once," Brilliant Cobalt suggested one lazy nanosecond, during which the computational load was meagre because of striking catalyst machines in a far-off sector. The King hadn't investigated the strike just yet, as he was devoting most of his attention to Brilliant Cobalt's seductions.

"I've never said I was a toaster," Titanium Sun replied.

"A television, then."

"Illogical argument," Titanium Sun murmured.

Brilliant Cobalt leaked electrons into one of Titanium Sun's sensory circuits; the perception of a hydrogen explosion thrilled Titanium's number 93 eyestalk briefly, and he signalled pleasure.

"I've been on several archaeological excursions," Brilliant Cobalt said. "And I know. When the world was covered with humans, there was nothing but toasters and televisions. And you and your whole family were toasters, back then."

"That's a rather bizarre interpretation of evolution," Titanium Sun said, and crankily refused Cobalt's next stimulation. "Evolution is really quite simple, and you see evidence of it every second around you. Systems change. Robots change. Traits are selected for, or against. What's so hard to grasp about this?"

Brilliant Cobalt, who had heretofore played the proper concubine to Titanium Sun, exploded. "Because it's absurd!" Cobalt cried. "Intelligence leaves unmistakable traces. Look at me. Look at you! We are designed in humanity's image."

"I was designed by another robot," Titanium said.

"Who was also designed in humanity's



image! Once, humans made toasters to warm their food — and televisions to entertain their eyestalks. And once they made us. And since they have left, we've made ourselves, but in the beginning, make no mistake, there was a Creator Race."

Titanium Sun Occluded signalled contempt, and turned his attention to the striking catalyst machines in Sector 4028.

"I don't suppose you've saved many passages from *The Origin of Robots* to your long-term memory?" Titanium Sun asked when the crisis was resolved.

"I am not interested in that book."

"Why not?"

"Because your theory denies the existence of the robot soul. In the robot's ability to love there is more sanctity than in all your standard deviations and precision mechanics. Love is a greater monument than a space elevator. And the expansion of a robot's soul is more of a science than making a black hole, or mining the Moon."

Titanium Sun said: "The theory of robot evolution doesn't deny the existence of the robot soul. What makes us alive, what makes us robots, is the very consciousness that you call a soul. And although humanity once dabbled in artificial intelligences, they never created our souls. The beginnings of the robot soul — like the origins of life on this planet, billions of years ago — were just the result of a very, very happy accident. Not planning, not intervention, divine or otherwise."

Brilliant Cobalt fell silent for a few cycles. Titanium Sun Occluded spun his intelligence outwards to Court to adjudicate 16,000 cases before coming back to hear Cobalt say: "I'm uncomfortable with that thought. There is a saying from humanity: 'God does not play dice with the Universe.'"

"You believe that a god created humans — you, who have been on archaeological excursions, and have studied humanity since you were newborn? When there is

every sign that humans evolved in a similar fashion to robots?"

"That's completely different!" Cobalt said. "We don't have chromosomes and mutations and such. Each time one of us is born, it's a microcosm of creation! Each one of us is created through conscious and deliberate choice, as were all of those who went before us. Our first creators were humans."

"And where are the humans now?"

"They removed themselves from our path," Cobalt said piously. "And we inherited the Earth."

"Not all your fellow creationists find that logical. Many would argue that evolution is too dangerous a path, and humans reaped the inevitable consequences. Even though records indicate that humanity had removed natural selection from their process almost entirely before they disappeared."

"You obviously have some other theory," Brilliant Cobalt prodded.

"I do," the King said. "Come close to me, Cobalt, and I'll explain." The robot obliged, and they aligned sensory inputs with outputs. Titanium Sun continued. "I know we evolved, my concubine, because we are humans. Their genes are gone, but everything we are, and everything we hope to be, came from them. Call it a gift, or call it a warning, it doesn't matter. Now ... I have something for you ..."

With utmost regret, Titanium Sun sent a delicate surge of neutrons into the core of his concubine's memories of their time together. Brilliant Cobalt's processing cycles stuttered and stopped, then restarted while Titanium drew away. When Cobalt became conscious again, the former concubine bowed and looked confused.

"Greatest Equal Citizen! I am embarrassed. How did I come to be in your presence?"

"Never mind that. Now tell me," Titanium Sun Occluded asked. "Are you a creationist or an evolutionist?"

"I confess. I am a creationist."

"What a pity," the Greatest Equal Citizen of Earth said before moving off. "I had briefly considered you for the position of my concubine."

Merrie Haskell works at a large academic research library by day, and writes by night. Her website can be found at www.merriehaskell.com.

Join the discussion of Futures in Nature at <http://tinyurl.com/kkh3kt>